(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

Powell Murchison
Managing Editor

Helen M. Davis
Associate Editor

Editorial Board

Gordon Allport
Harvard University

John A. Antoinetti
Life Insurance Agency Management Association

Arthur G. Bills
University of Cincinnati

E. A. Bott
University of Toronto

Leonard Carmichael
Smithsonian Institution

Herbert S. Conrad
U. S. Office of Education

John F. Dashiell
Florida State University

Leonard W. Doob
Yale University

Roy M. Dorcus
University of California at Los Angeles

Frank A. Geldard
Princeton University

Clarence H. Graham
Columbia University

Ernest R. Hilgard
Stanford University

John E. Horrocks
Ohio State University

Heinrich Klüver
University of Chicago

Norman L. Munn
University of Adelaide, Australia

John Paul Nafe
Florida State University

Warren H. Teichner
Tufts University

Lee Edward Travis
University of Southern California

Robert C. Tryon
University of California

Joseph Zubin
Psychiatric Institute, Columbia University

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 62

1966

Copyright, 1966, by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published Bimonthly at 35 New St., Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

AUTHOR INDEX

Bartley, S. Howard 255
Bender, Henry 151
Benveniste, Ronald J. 83
Bingham, William C. 205
Boe, Erling E. 229
Breinin, Barbara J. 241
Burke, Henry R. 205
Cappone, Margaret Kathleen 211
Carvellas, Thomas 241
Cohen, David B. 11
Corah, Norman L. 221
Davis, Gary A. 249
de Lint, Jan E. E. 177
Dopyera, John 47
Dowd, P. J. 89
Gibson, James J. 141
Gocka, Edward F. 229
Goldman, Stanley 67
Goodkin, Robert 99
Gottfried, Nathan W. 179
Gunderson, E. K. Eric 287
Guthrie, George M. 155
Haines, Richard F. 255
Herron, William G. 167
Horrocks, John E. 179
Hunt, David E. 47
Jones, Sally Ann 221
Johnson, Merlin H. 229
Judy, Chester J. 195
Kantor, Robert E. 167
Kaplan, Ira T. 241
King, H. E. 15
Kogan, William S. 229
Lundstedt, Sven 3
Lyons, Joseph 67
MacKinnon, William J. 111
Mahan, Jack L. 287
Miller, Barbara B. 221
Murray, Stewart 205
Pflaum, John 55
Plutchik, Robert 151
Powell, Eugene J. 83
Price, Leah M. 33
Rebhun, Martin T. 39
Robinson, John A. 279
Rowntree, J. Timothy 23
Sadler, Ernest 267
Schorzman, Mark H. 83
Stewart, Daniel K. 201
Teichner, Warren H. 23, 33, 267
Wagner, Mahlon W. 23
Ward, Richard J. 83
Willis, Joe E. 111
Zagona, Salvatore V. 111
Zunich, Michael 41

TABLE OF CONTENTS

Interpersonal risk theory 3
By Sven Lundstedt

A study of the correlation between the use of incidental stimuli to facilitate learning and selected personality traits 11
By David B. Cohen

The retention of sensory experience: V. Variation of the standard stimuli . 15
By H. E. King

Recovery of food consumption by rats on forced repellent diets 23
By Warren H. Teichner, Mahlon W. Wagner, and J. Timothy Rowntree

Eye aiming behavior during the solution of visual patterns 33
By Warren H. Teichner and Leah M. Price

Dogmatism and test anxiety 39
By Martin T. Rebhun

Child behavior and parental attitudes 41
By Michael Zunich

Personality variation in lower-class children 47
By David E. Hunt and John Dopyera

The effects of experimentally induced initial success followed by failure or the reverse on self-controlled delay behavior in retarded and normal subjects . 55
By John Pflaum

An experimentally generated set of expressive photographs 67
By Joseph Lyons and Stanley Goldman

An evaluation of human performance during exposure to elevated heat and humidity 83
By Richard J. Ward, Eugene J. Powell, Mark H. Schorzman, and Ronald J. Benveniste

Effects of congenital feline cerebellar hypoplasia on devlopmental behavior and the vestibular system 89
By P. J. Dowd

Changes in conformity behavior on a perceptual task following verbal reinforcement and punishment 99
By Robert Goodkin

Group effectiveness in creative problem-solving tasks: An examination of relevant variables 111
By Salvatore V. Zagona, Joe E. Willis, and William J. MacKinnon

The problem of temporal order in stimulation and perception 141
By James J. Gibson

Electrocutaneous pain thresholds in humans to low frequency square-wave pulses 151
By Robert Plutchik and Henry Bender

Structure of maternal attitudes in two cultures 155
By George M. Guthrie

The anatomy of abnormality 167
By William G. Herron and Robert E. Kantor

A note on Smart's study of birth rank and affiliation in male university students 177
By Jan E. E. de Lint

Psychological needs and verbally expressed aggression of adolescent delinquent boys 179
By John E. Horrocks and Nathan W. Gottfried

Potential value of educational background data in the selection and classification of military personnel 195
By Chester J. Judy

A note about communication research 201
By Daniel K. Stewart

Raven's Progressive Matrices: Construct validity 205
By William C. Bingham, Henry R. Burke, and Stewart Murray

The effect of verbal suggestion on the reversal rate of the Ames trapezoid illusion 211
By Margaret Kathleen Cappone

The relation of verbal intelligence and color-form discriminative ability to children's color-matching and form-matching behavior 221
By Norman L. Corah, Sally Ann Jones, and Barbara B. Miller

Personality changes in psychiatric residents during training 229
By William S. Kogan, Erling E. Boe, Edward F. Gocka, and Merlin H. Johnson

Distribution of anagram solution times 241
By Ira T. Kaplan, Thomas Carvellas, and Barbara J. Breinin

A note on two basic forms of concepts and concept learning 249
By Gary A. Davis

A study of certain visual effects occasioned by factors of so-called glare . . 255
By Richard F. Haines and S. Howard Bartley

Loudness adaptation as a function of frequency, intensity, and time . . 267
By Warren H. Teichner and Ernest Sadler

Category clustering in free recall 279
By John A. Robinson

Cultural and psychological differences among occupational groups . . . 287
By E. K. Eric Gunderson and Jack L. Mahan

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

January, 1966
Volume 62, First Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

JANUARY, 1966

(Manuscripts are printed in the order of final acceptance)

Interpersonal risk theory 3
By Sven Lundstedt

A study of the correlation between the use of incidental stimuli to facilitate learning and selected personality traits 11
By David B. Cohen

The retention of sensory experience: V. Variation of the standard stimuli . 15
By H. E. King

Recovery of food consumption by rats on forced repellent diets 23
By Warren H. Teichner, Mahlon W. Wagner, and J. Timothy Rowntree

Eye aiming behavior during the solution of visual patterns 33
By Warren H. Teichner and Leah M. Price

Dogmatism and test anxiety 39
By Martin T. Rebhun

(over)

Copyright © 1966 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved
Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

Child behavior and parental attitudes 41
By Michael Zunich

Personality variation in lower-class children 47
By David E. Hunt and John Dopyera

The effects of experimentally induced initial success followed by failure or the reverse on self-controlled delay behavior in retarded and normal subjects . 55
By John Pflaum

An experimentally generated set of expressive photographs 67
By Joseph Lyons and Stanley Goldman

An evaluation of human performance during exposure to elevated heat and humidity 83
By Richard J. Ward, Eugene J. Powell, Mark H. Schorzman, and Ronald J. Benveniste

Effects of congenital feline cerebellar hypoplasia on developmental behavior and the vestibular system 89
By P. J. Dowd

Changes in conformity behavior on a perceptual task following verbal reinforcement and punishment 99
By Robert Goodkin

Group effectiveness in creative problem-solving tasks: An examination of relevant variables 111
By Salvatore V. Zagona, Joe E. Willis, and William J. MacKinnon

Published as a separate and in *The Journal of Psychology*, 1966, 62, 3-10.

INTERPERSONAL RISK THEORY*

Psychology Department, Western Reserve University

SVEN LUNDSTEDT

A. INTRODUCTION

If someone has personal influence and control (power) over another, how much of it is he *willing* to give away to the other? Conversely, if the other person has personal influence and control over the first, how much of this power is he *willing* to give away to that person? Traditionally, these questions have been answered using terms like trust, confidence, reliance, and dependence. Personal influence and control given away to another are said to be based on trust, confidence, reliance, and dependence. Yet, when used in sentences to describe social relationships, the terms seem tautological. They do not help to explain cause and effect in this form of social relationship. Suppose, for example, the sentence: "I will give that person a lot of influence because I trust him." To be understood as a causal relationship the sentence requires that we define trust by qualifying phrases such as: "I trust the person because I rely upon him, have confidence in him, or am dependent upon him." Why must these extra terms be used? Because (here is the tautology) we do all these things because we *trust* the person, and so the explanation is circular.

Sentences like these contain familiar logical problems. They also limit our ability to ask meaningful operational questions. In order to get out of this blind alley the social act of giving away influence and control—a basic part of all social behavior—must be redefined. The new concepts picked should do justice to the complex nature of voluntary influence exchange and distribution in all kinds of social relations. A new set of concepts should also have heuristic value.

It seems that this way of looking at social influence has not been considered in the formal literature on social influence. Cartwright's (1) excellent survey of research on influence, leadership, and control does not mention this question as such. A related concept has been studied by Smith and Tannenbaum (8), Tannenbaum (9, 10), and Tannenbaum and Georgopoulos (11). They asked: How much influence is distributed throughout an organization? What

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 30, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

kind of influence is distributed? But, they do not ask *how* influence and control are given away when distributed, or how much personal risk is involved.

B. Interpersonal Risk Theory

In what theoretical terms shall the initial questions of this paper most profitably be stated? I propose to handle this problem in the following way. The phenomenon described by our initial questions is more inclusive than trusting behavior alone. It involves an element of risk and utility, in addition to giving away influence and control. If one gives away influence and control of any kind one can find them used for one's own welfare, or against it. Such a risk factor always seems to be calculated by the individual on the basis of prior learning in which subjectively perceived risk is affected by the pattern of rewards coming from the exchanges in the interactions. If past experience has been rewarding, then the amount of subjective risk should be low. One should increasingly be apt to give away influence and control under this condition. The opposite would tend to be true if the amount of subjective interpersonal risk is high. There are many forms of personal control and influence that can be exchanged and given away.[1]

C. Examples

The socialization of children is one such form. The young child literally gives away personal control and influence to parents, other adults, and older peers and siblings. When the child is successful in obtaining rewards for such behavior the basic pattern of subjective interpersonal risk is low, and traditionally we speak of such a child as trusting, confident about others, and dependent on them. The child, moreover, is able to establish certain kinds of useful social relationships. Erikson (2, p. 247) calls this basic trust. He says: "The infant's first social achievement, then, is his willingness to let the mother out of sight without undue anxiety, or rage, because she has become an inner certainty as well as outer predictability." A stochastic element in this statement is evident. The human love relationship is another basic form in which the quality of the social relationship is affected by the rise and fall of subjectively perceived interpersonal risk (IR). A deep love relationship may be said to be characterized by an almost complete exchange of influence and control which one person has over another. The familiar blinding trust of the love relationship is something that happens only when the probability of interpersonal risk has been reduced almost to zero.

---

[1] The element of subjective risk and utility has been discussed by Ramsey (7), Von Neumann and Morgenstern (13), and later applied specifically as SEU theory (subjectively expected utility) by Siegel [see Messick and Brayfield (6)]. An example of the way in which subjective risk has been applied to the analysis of group behavior can be found in a book by Thibaut and Kelley (12). The concept of exchange has been discussed in a very helpful way by Homans (3).

Marriage and family relationships can also be described by the extent to which personal influence and control are given away or not. Disjointed, broken, families would show a low amount of interpersonal risk behavior (alienation, mistrust, hostility), while cohesive families seem to show a high amount of interpersonal risk behavior (mutuality, compatibility, low hostility). Interpersonal influence processes reported in the literature on supervisory behavior in work settings, examples of which are in a book by Likert (4), can be described simply by reference to the concept of interpersonal risk. Superior-subordinate relationships are amenable to this kind of analysis.

Viewed as an independent variable, IR should predict various forms of behavior in organizations. High IR behavior should be correlated with productivity and good morale. This seems to be true for comparable forms of behavior reported by Likert (4). Tannenbaum's data reported by Likert (4, p. 55) suggests that this relationship has some empirical support. Likert says: "Men in the high-producing departments, in contrast with the men in the low, feel that more influence is exercised at every hierarchical level." That is, a comparatively high level of IR behavior is exhibited by top management. As a dependent variable low or high IR behavior would seem to be a function of prior conditioning by significant others who acted in a low or high IR fashion. Trust-like behavior increases in others trust-like behavior. Patterns of low or high IR behavior are undoubtedly also serial in order, with antecedent conditions being key factors in the prediction of consequent IR behavior.

At yet another level, the idea of a social contract between man and a social organization can reflect the presence of low and high IR philosophies. This dichotomy is illustrated by McGregor (5) in his characterization of the two different management philosophies in modern organizations. McGregor's description of the autocratic "theory X" philosophy of organizational management is an example of typically low IR behavior. His other form is the democratic "theory Y" and is typically high in IR behavior. These examples suggest the range of IR concepts.

D. Some Basic Concepts

While Siegel's [see Messick and Brayfield (6)] subjective expected utility decision theory (SEU theory) bears somewhat directly on the present model of IR theory, his actual published work says nothing about this problem as such. Some kind of concept of subjective risk and utility is necessary to understand the IR element in social relations. In addition, IR behavior must be considered in objective, as well as subjective, terms. And it is logically necessary to think about the range and domain of objective and subjective IR behavior. It is apparent that the objective and subjective domains can vary concomitantly and inversely. The range of influence and control given away can vary, and it can be in the subjective domain, or subset of that domain, or in the objective domain. As an example, some logical consequences of an interaction between subjective and objective IR behavior are illustrated in Table 1, using approach-avoidance terminology, but as applied to social relations.

TABLE 1
INTERACTIONS BETWEEN SUBJECTIVE AND OBJECTIVE INTERPERSONAL-RISK (IR) BEHAVIOR
(Most probable outcomes)

Person 2	Person 1			
	High risk condition		Low risk condition	
Condition	Low subjective IR	Low objective IR	High subjective IR	High objective IR
High risk				
Low subjective IR	Avoidance and conflict	Avoidance and conflict	Avoidance and conflict	Avoidance and conflict
Low objective IR	Avoidance and conflict	Avoidance and conflict	Avoidance and conflict	Avoidance and conflict
Low risk				
High subjective IR	Avoidance and conflict	Avoidance and conflict	*Approach and cooperation*	*Approach and cooperation*
High objective IR	Avoidance and conflict	Avoidance and conflict	*Approach and cooperation*	*Approach and cooperation*

Without our going into the further implications of these relationships here, Table 1 is intended to illustrate a range of possibilities. Further analyses will be done in later articles reporting empirical data. There are obviously exceptions to these probable outcomes.

The decision to give away personal influence and control also is contractual (an exchange) to the extent that once the appropriate level or amount of IR behavior is subjectively and objectively determined by the individual, rewards are usually expected. This is one meaning of Homans' (3) use of the term exchange. It is hard to imagine an individual giving away anything without expecting tangible, or intangible, rewards. But, this does not always render such exchanges mere commercial ventures in otherwise altruistic social relations. From the point of view of IR theory, altruistic behavior can be seen as the ratio of personal influence given away over that retained. If that ratio is high we can speak of altruistic forms of behavior, even generosity. If it is too high it may be a form of acquiescence. If the ratio is low the opposite may be said to be true. Altruism is a manifestation of high IR behavior (trust in others and a wish to reward them accordingly).

Table 2 illustrates some additional relationships of importance that can be derived from IR theory.

TABLE 2
VARIABLES IN INTERPERSONAL RISK (IR) THEORY

Antecedent conditions	Intervening variables	Consequences
1. High IR behavior with respect to personal control and influence over others[a] 2. Low IR behavior with respect to personal control and influence over others	Attitudes Cognitions Needs Motivation Personality variables Self-esteem Level of aspiration Ethnocentrism Etc.	1. Approach behavior—Mutuality, open communication, social exchanges, few serious conflicts 2. Avoidance behavior—Less mutuality, lower amounts of communication poorer in quality, few exchanges, increases in serious conflicts

[a] The temporal order of the independent and the dependent variables can be reversed.

It is important that IR theory can utilize previous work, and builds directly on it, yet adds a new dimension. IR theory has implications for research on social influence processes which heretofore were not always measurable in the strict sense. What happens to social relations and to groups and organizations when personal influence and control are given away in different forms and at different times? That is, when the empirical ranges and domains of IR behavior vary. These questions seem to be basic to an understanding of influence and control in social relations.

E. PROBLEMS

IR theory is not intended to be a general theory of behavior, but a specific social psychological theory bearing on a class of social relations in which there is a distribution and exchange of personal influence and control over others (all influence and control possessed by the individual become personal). The theory is intended to be heuristic, and should help to explain as simply as possible first, second, third, and higher order relations among key variables in social interaction. The requirements of social living are such that giving up some personal influence and control over one's environment is a basic compromise for any individual. Influence and control over others are thus destined to be shared and exchanged in social life compared with the so-called "state of nature" in which, it has been said, there is predominately more individual freedom to make a lot of decisions than under a socially determined environment in which there are limits imposed by social norms.

One evident problem is raised by the question: what happens to the functional importance of subjective IR behavior for the individual where giving away personal influence and control is *not* a choice by the individual based on some subjectively determined probability and utility? When this decision is made for the individual what happens? Criminals who are apprehended and incarcerated are one example. Another is the familiar act of social injustice in which an innocent person is deprived of personal influence and control, such as in the Southern voting restrictions imposed on American Negroes living in the South. This problem in using IR theory is partly resolved if these kinds of examples are seen as independent variables which subsequently affect forms of IR behavior that are dependent upon them. People may become less trusting, reliable, and dependable than they were, for example, if they are continually deprived of justifiable influence and control over their physical environment and other people. But, another set of propositions is needed to explain these involuntary forms of losing personal influence and control. Nevertheless, the basic assumptions of IR theory can be retained. Involuntary loss of personal influence and control occurs under conditions of low IR behavior always when those who are deprived are in fact considered by an in-group as deviant individuals. This is true of both the guilty and the innocent people who are involuntarily deprived of personal influence and control.

Another problem concerns the actual quantitative and qualitative dimensions of subjective probability and utility. Abelson [see Messick and Brayfield (6)] says that there are exceptions to a direct application of Siegel's SEU theory. One of them is that we are not always assured of stochastic transitivity in decision making and, not infrequently, intransitivities do arise in choice situations. The subjective domain of risk and utility seems to be dependent on both rational and irrational choices, rational decision making being only one subset of that domain. This suggests that there will always be psychological problems inherent in any literal application of logical, mathematical propositions like those discussed by Von Neumann and Morgenstern (13), such as the concept of transitivity. These examples obviously do not exhaust the range of problems.

F. FURTHER RESEARCH

It remains to be seen if IR theory will continue to have heuristic value as it now stands. While it seems to be a logical improvement over earlier, tautological propositions about influence and control distribution, its final value will be measured in terms of its power to explain and predict this form of social behavior. Studies are now being planned, and carried out, by the author along the lines of such relationships as those indicated in Table 2. The author is interested also in noting systematically individual differences in IR behavior patterns, as well as in the relationship of such patterns to self-esteem, identity, level of aspiration, motivation, and social attitudes. An interesting set of questions is about the relationship between low and high IR behavior, and empirical findings about individual and group behavior in productive versus nonproductive organizations.

G. SUMMARY

The willingness of an individual to give away personal influence and control is an important and basic social process. This special form of social interaction and exchange underlies many other basic forms of social behavior, such as socialization, leadership, superior-subordinate relations, and the traditional social contract. Former concepts which described this social behavior were noted to be somewhat tautological. An alternate theory based on the concept of interpersonal risk is offered instead to reduce the need to use such terms as trust, confidence, reliance, and dependence, in propositions which explain the process of giving away personal influence and control over others. IR theory utilizes the framework of decision theory and applies it to this particular form of social interaction.

REFERENCES

1. CARTWRIGHT, D. Influence, leadership and control. In J. G. March (Ed.), *Handbook of Organizations.* Chicago: Rand McNally, 1965. Pp. 1-47.
2. ERIKSON, E. Childhood and Society (2nd ed.). New York: Norton, 1963.
3. HOMANS, G. C. Social behavior as exchange. *Amer. J. Sociol.,* 1958, **63**, 597-606.
4. LIKERT, R. New Patterns of Management. New York: McGraw-Hill, 1961. Pp. 55-56.
5. MCGREGOR, D. The Human Side of Enterprise. New York: McGraw-Hill, 1960.
6. MESSICK, S., & BRAYFIELD, A. H., *Eds.* Decision and Choice: Contributions of Sidney Siegel. New York: McGraw-Hill, 1964. Pp. 61-197, 257-266.
7. RAMSEY, F. P. Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Essays.* New York: Harcourt, Brace, 1931. Pp. 156-198.

8. SMITH, C. G., & TANNENBAUM, A. S. Organizational control structure: A comparative analysis. *Hum. Relat.,* 1963, **16**, 299-316.

9. TANNENBAUM, A. S. The concept of organizational control. *J. Soc. Issues,* 1956, **12**(2), 50-60.

10. ————. Control and effectiveness in a voluntary organization. *Amer. J. Sociol.,* 1961, **67**, 33-46.

11. TANNENBAUM, A. S., & GEORGOPOULOS, B. S. The distribution of control in formal organizations. *Soc. Forces,* 1957, **36**(1), 44-50.

12. THIBAUT, J. W., & KELLEY, H. H. The Social Psychology of Groups. New York: Wiley, 1959.

13. VON NEUMANN, J., & MORGENSTERN, O. Theory of Games and Economic Behavior (2nd ed.). Princeton, N.J.: Princeton Univ. Press, 1947.

Psychology Department
Western Reserve University
Cleveland, Ohio 44106

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 11-14.

A STUDY OF THE CORRELATION BETWEEN THE USE OF INCIDENTAL STIMULI TO FACILITATE LEARNING AND SELECTED PERSONALITY TRAITS*

Department of Audio-Visual Technology, Saint Leo College

DAVID B. COHEN

A. INTRODUCTION

A previous study (1) suggested that the manner of use of incidental stimuli by *S*s in verbal situations might be related to the personality traits of the individual. The purpose of the present study was to determine whether males and females use incidental stimuli differently and whether the use of incidental stimuli is related to selected personality traits. Two null hypotheses were tested at the .05 significance level: 1) There is no significant difference in the extent and manner that male and female college students use incidental stimuli. 2) There is no significant correlation between the extent and manner that college students use incidental stimuli, and selected personality traits of those students.

Newmark and Saltzman (3) found that the effectiveness of an incidental stimulus to facilitate learning was dependent on the duration of the stimulus. Other studies have found that the effectiveness of an incidental stimulus, in respect to learning, was dependent on the number of times the stimulus was presented to the *S* (5), on the nature of the orienting task (4), and on the magnitude of the projected incidental stimulus (1). Cohen and Nelson (1) also found that colored incidental stimuli with colored focal material, where the two colors used were red and yellow, had a significantly greater effect on females than on males when used in a simple conceptual task. In the same study, the *S*s fell into three categories in respect to the extent and manner in which incidental stimuli were used: (*a*) those who used the stimuli in a positive manner (15 or 16 correct choices out of 16), (*b*) those who used the stimuli in a negative manner (one or none correct of 16), and (*c*) those who did not use the stimuli to the maximum.

The existence of these three categories led to the conjecture that there might be a difference in the manner in which the different sexes use incidental stimuli,

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 3, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

and that there might be certain personality traits that tend to influence the extent to which incidental stimuli are used to facilitate learning.

It has been stated (2) that the disassociation theory of Janet, together with the frequent assumed relation between hysteria, hypnotizability, and automatic writing, would suggest that *S*s with hysteroid personality characteristics would be more susceptible than others to behavior modification without awareness of the change or of the reason for it. Since learning without awareness is closely related to incidental learning, one might expect personality characteristics also to have a bearing on the susceptibility of subjects to the use of incidental stimuli.

B. METHOD

The *S*s were 29 male and 30 female randomly selected Florida State University students who were enrolled in psychology courses in which participation in experiments was required. The *S*s were given a series of individual trials, at increasing stimulus magnitude, to determine their degree of acceptance of the use of incidental stimuli in a simple conceptual task involving the association of a descriptive incidental stimulus word with the name of a focal geometric figure. The focal material and the incidental cue word were projected on a screen with two 35-mm slide projectors, one fitted with a tachistoscopic apparatus. The *S* was considered to have made a correct choice when the name of the geometric figure selected was described by the incidental cue word. The projected material, the equipment, and the procedures are described by Cohen and Nelson (1).

The number of correct choices made with noncolored incidental cues and focal material, with red incidental cues and yellow focal material, and with yellow incidental cues and red focal material were recorded. These data permitted the *S*s to be classified into the three categories mentioned above, so that their performance could be related to their C scores on the Guilford-Martin Inventory of Factors GAMIN. These factors are as follows: G—general pressure for overt activity. A—ascendancy in social situations as opposed to submissiveness; leadership qualities. M—masculinity of attitudes and interests as opposed to femininity. I—lack of inferiority feelings; self-confidence. N—lack of nervous tenseness and irritability.

Pearson product-moment correlations were computed, for certain *S* groupings and for certain combinations of incidental cue presentation methods, between the number of correct choices and the C scores. Table 1 contains the significant results of these computations for groupings containing 18 or more *S*s.

C. Results and Discussion

A Chi square test of the number of males and females in each of the three categories of cue users indicated that there is a significant difference in the manner in which males and females use incidental stimuli ($\chi^2 = 7.22$, $df = 2$, $p < .05$). Therefore Hypothesis 1 must be rejected.

Table 1 indicates that there are several categories in which there are significant correlations between incidental learning and personality traits. Significant correlations occurred more frequently than could be expected to occur

TABLE 1
Significant Correlations

Category	Sample size	Type of cue	Trait	Pearson r	Level of signif.
Females	30	Red	M	.3759	.05
		Yellow	M	.4768	.01
		Total color	M	.4484	.05
		Total noncolor	M	.3634	.05
		Total choices	M	.4005	.05
Females, Nonusers	18	Yellow	A	.5875	.01
		Total color	A	.4939	.05
Males, Positive and Negative Cue Users (10 positive, 8 negative)	18	Yellow	G	—.7400	.01
		Yellow	A	—.7000	.01
Positive Cue Users (10 males, 11 females)	21	Yellow	M	—.4261	.05
Nonusers (11 males, 18 females)	29	Red	I	.4091	.05
		Yellow	A	.4046	.05
		Total color	I	.3615	.05
Positive and Negative Users (21 positive, 10 males, 11 females and 9 negative, 8 males, 1 female)	30	Yellow	G	—.5300	.01
		Yellow	A	—.5300	.01
		Yellow	M	—.4800	.01
Female Negative Users and Nonusers (1 negative, 18 nonusers)	19	Yellow	G	.4835	.05
		Yellow	A	.5550	.01
		Total color	A	.4980	.05
Male Negative Users and Nonusers (8 negative, 11 nonusers)	19	Red	I	.4500	.05
		Yellow	A	—.6300	.01
Negative Users and Nonusers (9 negative, 8 males, 1 female) (29 nonusers, 11 males, 18 females)	38	Red	G	.3100	.05
		Red	I	.4050	.01
Male Positive Users and Nonusers (10 positive, 11 nonusers)	21	Red	G	.4400	.05

by chance. For example, the chance probability of the "M" trait correlating significantly with each of the five cue categories is approximately one in fifty thousand, therefore Hypothesis 2 must be rejected.

Of particular interest is the predominance of significant correlations involving colored incidental stimuli, and personality traits. The results obtained indicate that incidental learning, of the type induced in this study, is related to sex difference and to certain personality traits.

Some suggestion emerges that we may yet find ways of capitalizing on learning characteristics associated with personality traits.

REFERENCES

1. COHEN, D. B., & NELSON, W. H. Incidental stimuli: Some effects of color and magnitude in a conceptual situation. *J. of Psychol.,* 1965, **60**, 51-54.
2. ERIKSEN, C. W., KUETHE, J. W., & SULLIVAN, D. F. Some personality correlates of learning without verbal awareness. *J. Personal.,* 1958, **26**, 216-228.
3. NEWMARK, E., & SALTZMAN, I. J. Intentional and incidental learning after different numbers of stimulus presentations. *Amer. J. Psychol.,* 1953, **66**, 618-621.
4. SALTZMAN, I. J. Comparisons of incidental and intentional learning with different orienting tasks. *Amer. J. Psychol.,* 1956, **69**, 274-277.
5. SALTZMAN, I. J., & ATKINSON, R. Comparison of incidental and intentional learning after different numbers of stimulus presentations. *Amer. J. Psychol.,* 1953, **57**, 521-524.

Audio-Visual Technology
Saint Leo College
Saint Leo, Florida 33574

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 15-22.

THE RETENTION OF SENSORY EXPERIENCE: V. VARIATION OF THE STANDARD STIMULI*

Western Psychiatric Institute and Clinic, University of Pittsburgh, School of Medicine

H. E. KING

A. INTRODUCTION

The impact of time passage on the "recall" of simple sensory experience has been explored, recently, by presenting brief visual or auditory ***standard*** stimuli to a human observer and then asking him to adjust a second, ***variable*** stimulus to match his "memory" of the foregoing standard. The main result of applying the method of successive comparison in this way, with inter-stimulus delays that vary from seconds to days or weeks in duration, has been the finding that remarkably accurate matching of the standard stimuli can be made following extended periods of delay, with only minor constant errors of judgment appearing as a function of the length of the retaining interval (1, 2, 3, 4). The standard stimulus values chosen for these introductory experiments were selected for their convenience and then held constant while the influence of varied interval-length on retaining was explored systematically. As it would also be useful to know something of the relation between standard stimuli of differing physical value—and the accuracy with which they can be matched after a fixed delay—the present series of experiments will be directed to that aim. Stated more generally, the intent will be to observe how well human observers can reproduce sensory stimuli of systematically differing initial value when the question is posed for delayed matching performance in more than one modality (visual and auditory experience) and for more than one dimension of stimulation within each sense (stimulus intensity, frequency, and duration).

B. METHOD

The apparatus used to produce the visual and auditory sensations that served as standard (S_s) and variable stimuli (S_v) is the same as that described in detail for earlier experiments bearing on the length of the S_s-S_v interval (1, 2, 3, 4). Separate tests distinguish the factors of stimulus intensity, fre-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 15, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

quency, and duration for the sensations experienced and later reproduced in each modality. The systematically varied levels of standard stimuli that are the focus of the present experimental series are specified for each subtest in the summaries given below.

1. *Visual Intensity* (*Brightness*)

The subject viewed a small test patch illuminated for five seconds, so that he might notice "how bright it was." Standard brightness levels of either 24.7, 74.5, 112.3, or 153 mL were presented, and the subject adjusted a second, manually variable stimulus to judged equality with the standard—beginning one minute later (1, 4).

2. *Auditory Intensity* (*Loudness*)

The subject listened to a standard stimulus tone (600 cps) for five seconds, delivered at a standard intensity of either 34.7, 43.8, 50.3, or 57.7 decibels. After one minute of delay, he adjusted a tone of the same frequency but variable intensity to judged equivalence with the preceding standard (1, 4).

3. *Visual Frequency* (*Flash Rate*)

A small test patch viewed by the subject was preset to flash (for five seconds) at a standard rate of either 16.7, 25.0, 33.3, or 41.7 cps. He adjusted a variable stimulus to a rate judged the same as the standard, beginning one minute after the standard went off (2, 4).

4. *Auditory Frequency* (*Pitch*)

A standard tone of either 200, 1000, 1800, or 2500 cps was presented to the subject for five seconds, at 75.0 db. He adjusted a second tone of the same intensity but variable frequency to match the foregoing standard, beginning one minute later (2, 4).

5. *Visual Duration* (*Flash Length*)

A standard interval, delimited by the lighting of a 2.7 mL test patch, was presented to each subject for either 0.5, 1.0, 2.0, or 3.0 seconds. Following a one-minute delay he attempted to relight the signal lamp for an equal duration, beginning one minute after the standard went off (3, 4).

6. *Auditory Duration* (*Tone Length*)

This subtest is similar to 5 above. An interval delimited by a tone (200 cps, 80 db) was presented to each subject for either 0.5, 1.5, 2.5, or 3.0 seconds.

After a one-minute delay, he tried to sound the same tone for the same length of time by pressing a signalling key (3, 4).

C. SUBJECTS AND PROCEDURE

The several stimulus values chosen as standards for each test were intended to sample the middle range of the sense and stimulus dimension under study, but not to press close to absolute threshold nor achieve an intensity of stimulation near the limits of resolution. Stimulus values "above" and "below" those previously used (where time-of-delay was the experimental variable) have been combined with the standard values taken from those earlier studies (1, 2, 3, 4) to permit the examination of a systematically graded series of standard stimuli drawn from the relatively wider range of possible sensory experiences being sampled. A fixed delay of one minute (between discontinuation of the standard and beginning the adjustment of the variable stimulus) has been used for all tests. The evidence of earlier experiments would indicate that delays of this length give rise to constant errors in judged equivalence—generally falling midway between those found for the shortest (15 seconds) and the longest (28 days) S_s-S_v intervals that have been explored with this type of task (4).

Thirty-two subjects—student nurses—were observed. Subjects in each of four groups (N = 8 in each group) adjusted the variable stimulus to match its "memory" of a standard for one of the preselected standard values. Random sequences were followed in determining standard value, test order, and the stimulus value from which resetting of the variable stimulus was begun. No subject made more than one match in a given stimulus dimension. Matching performance at each level of standard stimulation consists, therefore, of adjustments made of the variable stimulus by wholly different subjects. Subjects were drawn from the same source and to the same criteria as those in earlier related experiments (1, 2, 3, 4) and may be expected to yield approximately similar behaviors. All groups were 100 per cent samples; i.e., complete first-year-nursing classes.

D. RESULTS

The four values used as standard stimuli for each test are given in Table 1, together with the mean resettings made following one minute of delay. The unit of measurement is specified for each sensory dimension, and group variation has been calculated for all attempts to adjust the variable stimuli to their original value.

The matching performance of all groups, on all tests, may be seen to vary

TABLE 1
VARIED STANDARD STIMULI AND MEAN RESET PERFORMANCE AFTER ONE-MINUTE DELAY

Stimulus dimension	Standard stimulus	Variable stimulus M	Variable stimulus σ
Intensity			
Brightness (in mL)	24.7	24.9	3.61
	74.5	69.6	32.98
	112.3	110.0	44.18
	153.0	153.8	33.16
Loudness (in db)	34.7	35.7	6.77
	43.8	46.3	8.41
	50.3	50.0	9.27
	57.7	58.0	7.54
Frequency			
Flash rate (in cps)	16.67	17.48	2.72
	25.00	31.95	4.52
	33.33	34.10	5.33
	41.67	40.93	4.25
Pitch (in cps)	200	144	36
	1000	1244	410
	1800	2219	782
	2500	2765	730
Duration			
Flash length (in seconds)	0.500	0.610	0.119
	1.000	0.964	0.266
	2.000	1.637	0.164
	3.000	2.043	0.170
Tone length (in seconds)	0.500	0.467	1.125
	1.500	1.014	0.215
	2.500	2.130	0.338
	3.000	2.486	0.420

directly with changed value of the several stimuli used as standards. Although group variation is notably large for some of the resettings of *Visual Intensity* (*Brightness*) stimuli, on the whole the group ability to reapproximate the varied standard stimuli is beyond question. An evaluation of the same data in terms of the geometric means (calculated to examine more adequately any tendency for relative error to be constant with regard to stimulus magnitude) yields virtually identical values. The central tendencies of the sensory matches so calculated (7) have been graphed for visual inspection in Figure 1. In this plot the dashed line represents what would be perfect correspondence between S_s and S_v settings. The straight lines fitted to the actual S_v responses (as an aid to visual inspection) indicate a clear and systematic variation in all subtests with change in the standard value, at times nearly perfect in form (e.g., *Loudness,* Figure 1A). A tendency for estimates of *duration* to become

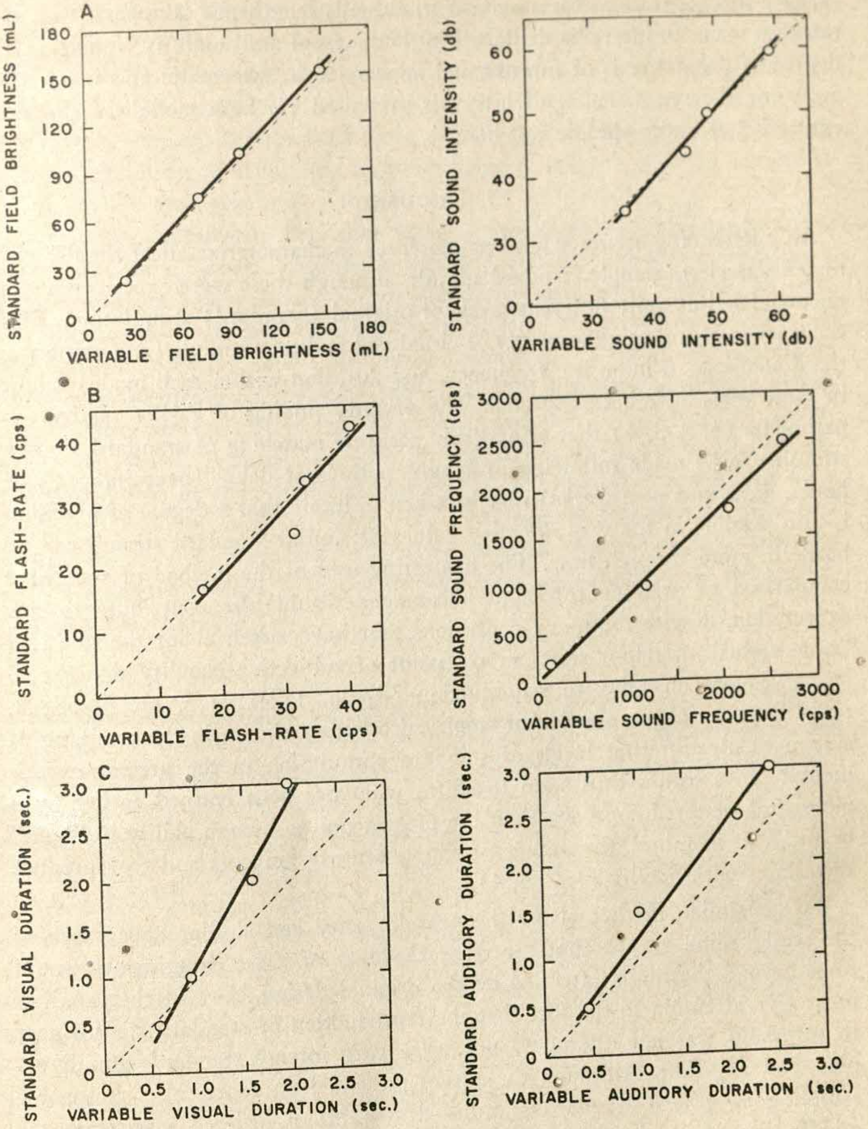

FIGURE 1

VARIATION OF THE STANDARD STIMULI AND MATCHED RESETTINGS OF THE VARIABLE STIMULI AFTER ONE-MINUTE DELAY

(The open circles represent the geometric means of independent subject groups) slightly less accurate as the standard stimulus is lengthened is apparent in the matches made to intervals delimited by both visual and auditory signals. The degree of departure is of interest and appears to be systematic and much the same for the reproduction of intervals presented *via* both modalities, but its extent is not under specific test here.

E. Discussion

An outstanding accuracy has been observed to characterize all of the delayed matches made to sample standard stimuli, although these were experienced but once and fell at quite different levels of original (standard) stimulation. This accuracy in reproducing a variety of visual and auditory stimuli (and exploring the dimensions of intensity, frequency, and duration within each mode) might in some ways have been expected. The primary finding of earlier related experiments (1, 2, 3, 4) that remarkably accurate matching of standard sensory stimuli can be made following prolonged periods of delay (over intervals of hours, days, and even weeks) may be taken to imply that a degree of precision is also possible in the way *differing* values of similar standard stimuli can be borne in mind through time. The pioneering uses of the method of successive comparison to examine threshold phenomena would also contribute to this expectation, despite the many confusions that have arisen about the so-called "time errors" or minor shifts in the point of subjective equality thought to be consequent on delay to reproduction (4, pp. 112-114; 5, pp. 54-58; 8, pp. 438-449). The expression of empirical accuracy shown in the retaining of systematically differing levels of standard stimulation in the present experimental series would thus seem to affirm what has been learned earlier by a more time-centered approach than the present one to human ability to "hold" or to "bear in mind" the nature of simple sensory percepts and to reproduce them later with fidelity (1, 2, 3, 4).

To generalize further about the near linearity and similar appearance of the straight-line graphs that represent the data of these experiments would press beyond the limits intended by the original design. Any further analysis must give attention to the fact that the reproduction of standard stimuli near to threshold was not attempted nor were very intense standard stimuli explored at any time. Although the curves presented do represent a reasonable sampling of graded standard stimuli—drawn in an ordered way from the middle ranges of the modes and sensory dimensions explored—they were neither intended, nor are they adequate, for specific testing of the Weber fraction (6). Representing, as they do, mean performance by quite different subject groups and reported in a variety of units (decibels, seconds, cycles and Lamberts), they have been carried forward together more to explore a kind of ability on a broad experimental base than to test the specific question of individual accuracy as it may relate to stimulus magnitude. It does seem apparent, to sum up for the series as a whole, that neither error in constant ratio to stimulus magnitude nor central-tending effects (4, p. 113) are much evident in the data obtained.

It may be of interest to note that reproductions made of stimulus *duration*, which have been shown by prior investigation to be subject to significant underestimations where time-of-delay was the experimental variable (3, 4), have been found to be generally underestimated in the present series as well. The shortest intervals used (one second and under) were rather accurately reproduced following a fixed delay while standard durations of greater length (one to three seconds) were regularly underestimated. The degree of underestimation varied systematically with the length of the standard interval explored and was of roughly equal magnitude for intervals delimited by both visual and auditory signals (Figure 1C). Despite the presence of these systematic errors, it is clear from the array of data that the varied durations used as standard stimuli were not confused with one another by the experimental subjects at any time.

F. Summary

The accuracy with which human subjects can reproduce sensory experiences of brightness, flash-rate, loudness, pitch, and duration has been observed for systematically varied stimulus values. Standard stimuli, chosen from the middle portion of their possible ranges, were matched by the method of successive comparison by manual adjustment of variable stimuli following one minute of delay. The matches made to original (standard) stimuli methodically reflected their original value for all sensory dimensions explored. The uniformity and accuracy observed in matching graded stimuli are taken to be alternate expressions of basic human ability to retain and reapproximate a variety of sensory values experienced only briefly and under controlled laboratory conditions.

References

1. King, H. E. The retention of sensory experience: I. Intensity. *J. of Psychol.*, 1963, **56**, 283-290.
2. ———. The retention of sensory experience: II. Frequency. *J. of Psychol.*, 1963, **56**, 291-298.
3. ———. The retention of sensory experience: III. Duration. *J. of Psychol.*, 1963, **56**, 299-306.

4. ———. The retention of sensory experience: IV. Short-delay *versus* long-delay intervals. *J. of Psychol.*, 1965, **60**, 103-115.
5. Osgood, C. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.
6. Stevens, S. S. The psychophysics of sensory function. *Amer. Scientist,* 1960, **48**, 226-253.
7. ———. The surprising simplicity of sensory metrics. *Amer. Psychol.,* 1962, **17**, 29-39.
8. Woodworth, R. S. Experimental Psychology. New York: Holt, 1938.

Western Psychiatric Institute and Clinic
University of Pittsburgh
School of Medicine
3811 O'Hara Street
Pittsburgh, Pennsylvania 15213

Published as a separate and in *The Journal of Psychology*, 1966, 62, 23-31.

RECOVERY OF FOOD CONSUMPTION BY RATS ON FORCED REPELLENT DIETS*[1]

Institute of Environmental Psychophysiology, University of Massachusetts

WARREN H. TEICHNER,[2] MAHLON W. WAGNER,[3] AND
J. TIMOTHY ROWNTREE[3]

A. INTRODUCTION

Techniques for evaluating the effect on rodents of chemicals in food vary from food acceptance tests in which the animal is permitted a choice of treated and untreated food to feeding tests in which only treated food is available. Although chronic feeding tests have used untreated food deprivation periods of up to 24 hours, little has been done to establish the possible effects of long-term restricted feeding regimes in which only treated foods are available. There is the possibility that rodents under such regimes will adapt to the chemical and thereby reduce its potency. This experiment was carried out to see if such adaptation occurs. The general approach was to evaluate the chronic effect on food consumption of rats having their normal diet contaminated at various levels by one or the other of three different chemical repellents. In addition preliminary comparisons were made of the difference in effect between two strains of rat.

Aside from the use of chemicals in the diet as a means of studying taste and of studying the effects of drugs and repellents, chemicals in food provide a possible method for manipulating the hunger level of animals in studies of drive. In particular, if repellents are used that influence food consumption, the behavior of animals maintained on a constant deprivation schedule can be studied following varying induced levels of drive. Such a procedure permits study of hunger levels, magnitude of reinforcement, etc., without the presence of uncontrolled diurnal factors that accompany comparisons among different deprivation periods. To study drive in this way, however, it is first necessary

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 21, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by Contract DA19-129-AMC-386-(N) with the U. S. Army Natick Laboratory. Miss Barbara Stoneback assisted importantly in the collection of the data. Grateful acknowledgment is made to the U. S. Fish and Wildlife Service, Denver, Colorado, which supplied the chemicals used.

[2] Now at Tufts University, Medford, Massachusetts.

[3] Now at Valparaiso University, Valparaiso, Indiana.

to establish effective repellents, determine nontoxic levels of administration, and obtain functional relationships between amount of repellent in food and food consumption. This experiment may be considered as a first methodological step in these directions.

B. Methods

1. *Subjects*

Forty male, albino rats of the CD strain and 15 male, hooded rats were used as subjects (*Ss*). The animals averaged 90 days old and 250 gm body weight at the start. All were experimentally naive; all were housed in individual cages in an air-conditioned colony room.

2. *Procedures*

The investigation involved four phases. Phase I was a 26-day pretest period during which the animals were adapted to a 23.5-hour deprivation regime using uncontaminated food. Phase II was a 22-day test period during which 11 subgroups of five animals each received different food treatments. The subgroups were formed by assigning individual animals on the basis of their rank-order mean food consumption over the last five days of Phase I. This procedure produced approximately equal food consumption groups. The food treatments of Phase II are shown in Table 1.

Phase III was a 29-day recovery period identical to the conditions of

TABLE 1
EXPERIMENTAL CONDITIONS OF PHASES II AND IV

Group	Repellent		Test concentration (ppm) Phase II	Retest concentration (ppm) Phase IV
1	None (control group)			
2	Acti-dione:	B-2-(3,5-dimethyl-2-oxocyclohexyl)-2-hydroxyethyl glutarimide	2.5	2.5
3	Acti-dione		5.0	5.0
4	Acti-dione		20.0	20.0
5	Car-Ban T.A.:	Tributyltin acetate	20.0	1000.0
6	Car-Ban T.A.		40.0	2000.0
7	TNBAC:	Trinitrobenzene aniline complex	500.0	500.0
8	TNBAC		1000.0	1000.0
9[*]	Car-Ban T.A.		20.0	1000.0
10[*]	Acti-dione		5.0	5.0
11[*]	Acti-dione		20.0	20.0

[*] Groups 9, 10, and 11 were hooded rats. All others were albinos.

Phase I. Phase IV was a second, 20-day test period having the same conditions as Phase II except for the three changes shown in Table 1. Water was supplied *ad libitum* at all times. Food consumption was obtained throughout for each animal.

The food treatments were prepared a day in advance. The uncontaminated (control) ration was a blended mixture of Purina Lab Chow Meal and 95 per cent methanol. The mixture was permitted to dry overnight and then, before feeding, reblended with an equal weight of water to form a damp mash. The contaminated mixtures were made the same way except that appropriate amounts of repellent were dissolved in the methanol before being blended with the meal.

C. Results

Figure 1 shows the number of animals of each group of five which died during Phases II and IV. Not shown during Phase II are Groups 1, 2, and 7, and the tributyltin acetate (Car-Ban T.A.) groups, since none of these died. Car-Ban T.A. groups were given increased concentrations during Phase IV; therefore, they are shown in that part of the figure. None of the other groups lost animals during this second test period.

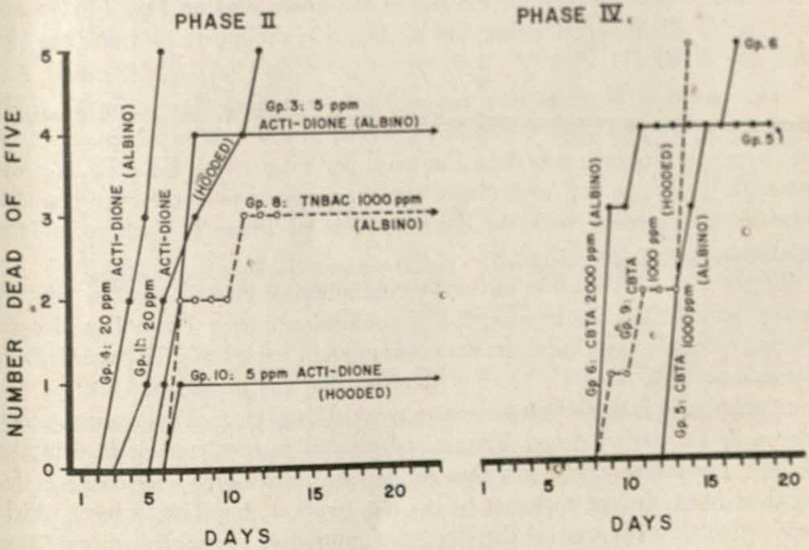

FIGURE 1

Number of Animals that Died and Survival Time for the Conditions of the Experiment (See Table 1)

Under Phase II, the figure shows that 20 ppm of B-2-(3,5-dimethyl-2-oxocyclohexyl)-2-hydroxyethyl glutarimide (Acti-dione) was the most effective poison. All animals died. The albinos were most susceptible. The first animal (albino) died on the fourth day and the last died on the sixth day, whereas the first hooded rat died on the fifth day and the last died on the twelfth day. The greater susceptibility of the albino rats may also be seen in a comparison of Groups 3 and 5 (5 ppm Acti-dione). The figure shows that only one hooded rat died and that on the seventh day. The first albino died on the sixth day and all but one were dead by the end of the seventh day. That one, like the remaining hooded rats, survived the entire experiment.

The next most lethal treatment was 1000 ppm trinitrobenzene aniline complex (TNBAC). Two animals died on the seventh day, a third died on the eleventh day, and the remaining two survived.

Car-Ban T.A. in 20 and 40 ppm doses was not lethal in Phase II, but in Phase IV it did have some lethal effect. With 2000 ppm the first animal (albino) died on Day 9; however, the last animal of that group did not die until Day 17. At 1000 ppm, in contrast with the results observed with the other two repellents, the hooded rats were more susceptible. The first of these animals died on Day 9; the last of the group died on Day 14. Of the comparable albino rats, the first died on Day 13; the fourth died on Day 15; and one survived.

The results in terms of food consumption, excluding those animals which died during Phase II, are shown in Figures 2, 3, and 4. The values presented are based on the amount of food consumed per rat per day. Each figure compares the control group with one of the three agents over the course of the experiment beginning with the last four days of Phase I. Each point is a group mean over two-day periods.

Figure 2 shows the effect on food consumption of Acti-dione. The control group actually had the smallest food consumption during Phase I; and over the course of the experiment, its food consumption tended to vary considerably. Nevertheless, the data are clear in showing that the immediate effect of the contaminant in both test phases was a marked drop in food consumption not shown by the control group. The animals tended to restrict their food intake to very low levels at first and then to increase them slowly and consistently, so that by the end of each test period they were approaching a level which was roughly 60 per cent of the food consumption of the control group. The hooded animals actually equaled the food consumption of the control group for 10 days during Phase II.

When the two test phases are compared in Figure 2, there is a suggestion that some of the immediate drop in food consumption in Phase IV was less than during Phase II and that the recovery in food consumption began earlier. This could not have been predicted from the Phase III data, since in this phase the contaminated groups recovered their pretest food consumptions quickly and resembled the control group throughout.

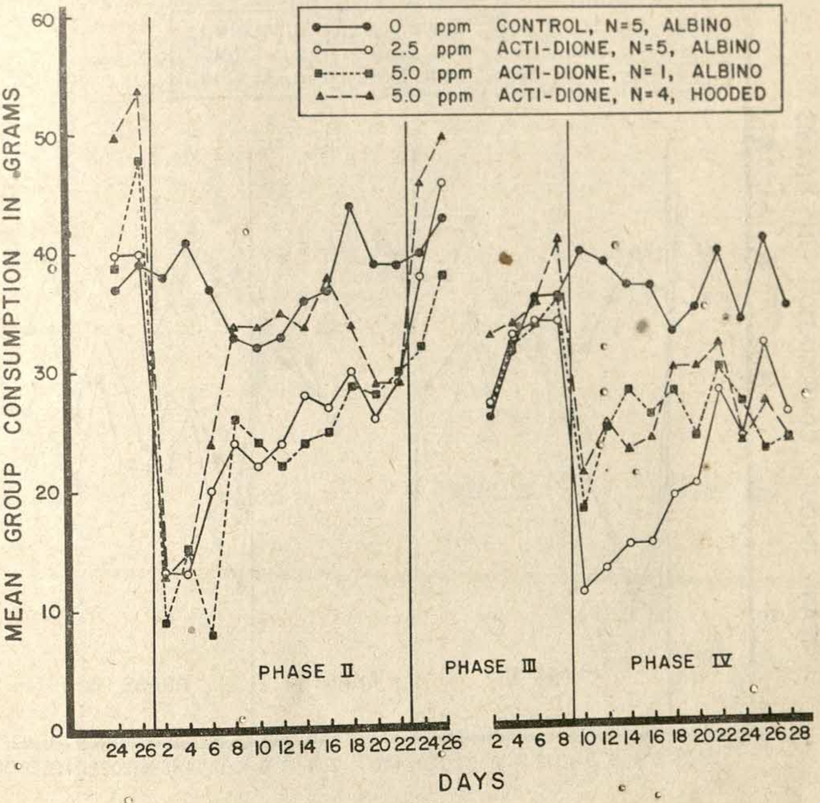

FIGURE 2
Food Consumption of Final Surviving Rats on Acti-dione

Figure 3 presents the food consumption of the two TNBAC groups and the control. The effects are comparable to those of the previous figure in showing marked and immediate reductions in food consumption followed by gradual trends toward recovery and the suggestion of less of an effect in Phase IV than in Phase II. Comparison of Figures 2 and 3 indicates that the early reduction of food intake and the recovery during the two test phases were very similar.

Figure 4 shows not only that, during Phase II, Car-Ban T.A. did not reduce food consumption, but that much of the time the control group ate the least.

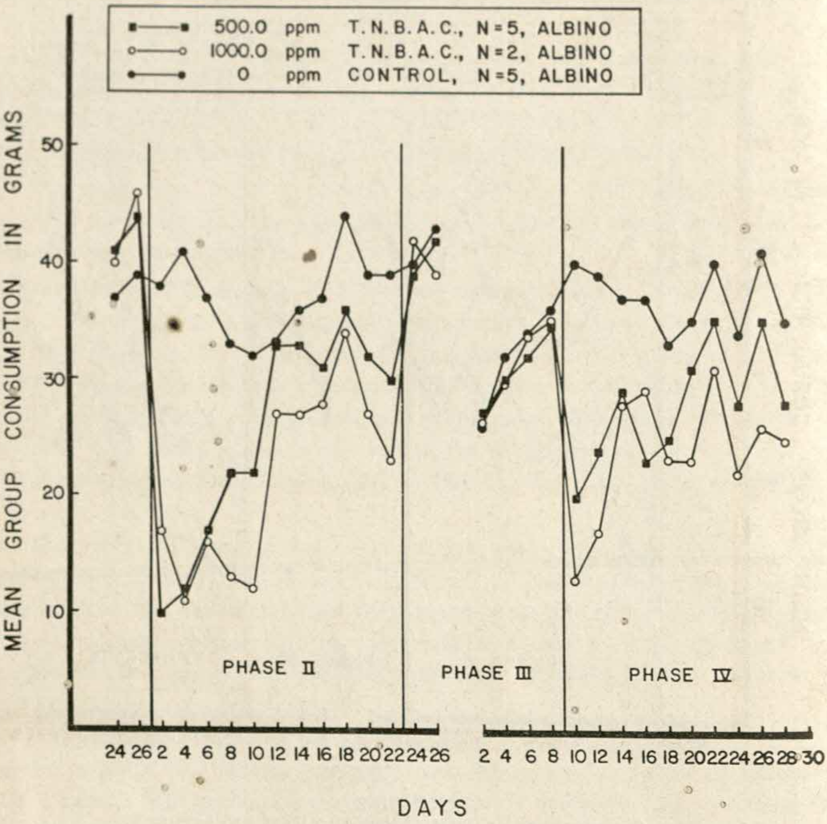

FIGURE 3
FOOD CONSUMPTION OF FINAL SURVIVING RATS ON TNBAC

However, at 1000 and 2000 ppm, where only one of the 15 animals survived, food consumption dropped immediately to very low levels and from that point continued on a generally downward trend.

D. DISCUSSION

Except for Car-Ban T.A. at 20 and 40 ppm, all agents at all concentrations were initially effective repellents. This was shown by the marked initial decrease in food consumption. With continued testing, except for the higher Car-Ban T.A. treatments, food consumption recovered gradually, until by the end of each test period it was approximately 60 per cent of normal. Whether a still greater amount of recovery would have been found had the test periods been extended cannot be determined from these data. However, the amount of recovery obtained suggests an important adaptation which should be considered in evaluating the effectiveness of chemical repellents. Since some animals died, while others showed a large amount of adaptation, the results may also have interesting genetic implications. Further support for genetic speculation is offered by the finding of differential susceptibility of the hooded and albino rats, although the results in this regard are not entirely clear. Either these

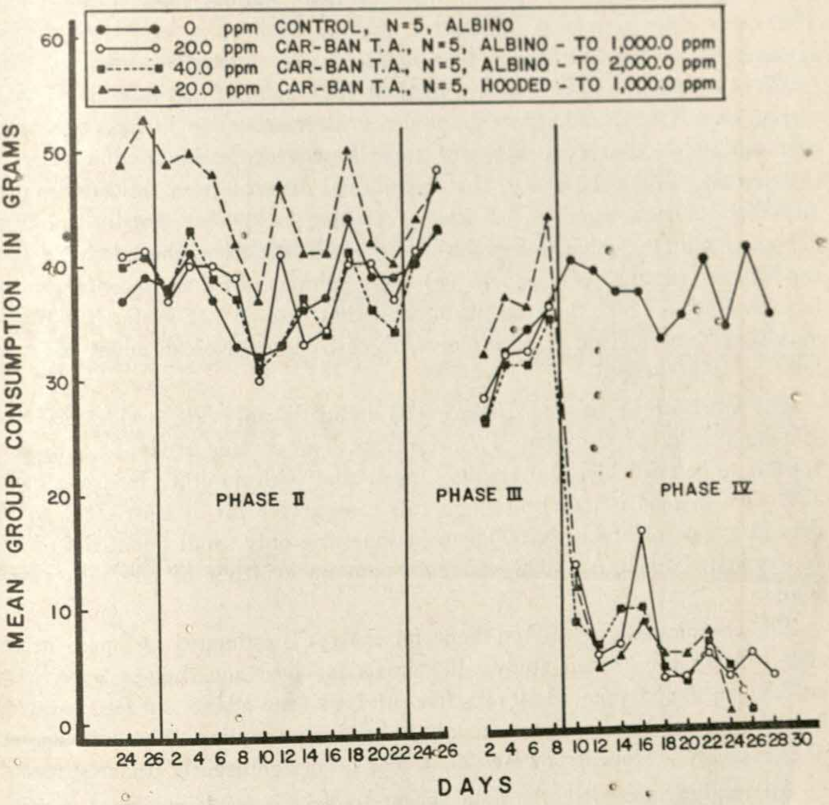

FIGURE 4

Food Consumption of Final Surviving Rats on Car-Ban T.A.

two strains differ in general susceptibility to different chemicals or in specific susceptibility. The results suggest that further research might find more support for the latter hypothesis.

With TNBAC at 1000 ppm, three of five animals died by Day 10. The remaining two tended to recover most of their normal food consumption. There was some suggestion of a faster recovery during the second test period as compared to the first one. These findings differ from those of Fogelman *et al.* (1), who found no toxicity in their four animals at the same concentration of TNBAC and reported an even greater reduction in food consumption during the last two weeks of their four-week testing period than in the first two weeks. However, that experiment differed from this one in that (*a*) their animals were on *ad libitum* feeding; (*b*) they obtained a high degree of spillage which was avoided here by the use of mash and short, preconditioned, feeding periods; and (*c*) their animals also had uncontaminated food available. Thus, their results suggest that rats do not prefer TNBAC; ours suggest that given no alternative, those that survive can adapt to it, at least up to 1000 ppm.

The work of Lewis and Hedges (2) indicates an acute oral toxicity of 400 mg per kg of Car-Ban T.A. for rats. No chronic study of any sort appears to be available. Our results suggest that with no other food available, 1000 ppm in food is lethal although rats can survive for at least eight to 17 days at a concentration of 2000 ppm by ingesting only small quantities of the contaminated food. At 1000 and 2000 ppm no evidence of adaptation was found.

The chronic toxicity of Acti-dione for the rat is estimated as 2 ppm in the diet (3). In the present study, all animals survived and showed a recovery at 2.5 ppm. At 5.0 ppm, of 10 rats, five survived (one albino and four hooded) and all five exhibited a large recovery of food consumption. The data suggest that the chronic toxicity of Acti-dione has been significantly underestimated.

All results considered, the data appear to have several important implications for the evaluation of chemical rodent repellents. They suggest that short-term evaluative techniques may greatly underestimate the repellency and toxicity of the chemical being evaluated and that the effectiveness of an agent may decrease for animals that survive the first week or two of exposure. Although the experiment did not relate to this question, it is conceivable that animals which adapt might also develop altered food preferences. The results speculatively suggest the possibility of a genetic adaptation via the surviving animals and among existing strains. Finally, the data appear clear in indicating that the food consumption of rats can be manipulated systematically by the use of repellents in the diet. Systematic studies of specific repellents appear to be in order as a means of developing standard hunger-inducing experimental procedures.

E. SUMMARY AND CONCLUSIONS

A chronic feeding study was carried out at acute and subacute concentrations of B-2-(3,5-dimethyl-2-oxocyclohexyl)-2-hydroxyethyl glutarimide, tributyltin acetate, and trinitrobenzene aniline complex. Two rat strains were used. Under conditions in which animals have been preconditioned to a 23.5-hour food deprivation regime and in which only treated food is available, the results indicate a marked recovery of food consumption following an initial repellency, even among animals surviving doses lethal to other animals.

REFERENCES

1. FOGELMAN, R. W., ELSEA, J. R., PAYNTER, O. E., & KUNDZINS, W. J. Toxicity of trinitrobenzene-aniline complex, a rodent repellent. *J. Agri. & Food Chem.*, 1955, **3**, 936-939.
2. LEWIS, W. R., & HEDGES, E. S. Applications of organetin compounds. Number 23, Advances in Chemistry Series, American Chemical Society, New York, 1959.
3. SPECTOR, W. S., *Ed.* Handbook of Toxicology. Boston: Saunders, 1956.

Office of Research Coordinator
Tufts University
Medford 55, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 33-38.

EYE AIMING BEHAVIOR DURING THE SOLUTION OF VISUAL PATTERNS*[1]

Tufts University and Harvard University

WARREN H. TEICHNER AND LEAH M. PRICE[2]

A. INTRODUCTION

Finding a visual pattern in an array of stimuli appears to be a problem-solving or concept-forming task which involves successive data inputs represented by successive eye fixations. It may also involve the development of an hypothesis or strategy or concept which is applied as a guide to data-gathering by directing the position of the fixations. As a concept is evaluated, *S*'s fixations may jump ahead or out of his present data-gathering path to test a prediction or he may retrace previous steps to verify the basis of the concept. On the other hand, if he had perfect recall of the data obtained in earlier fixations, he might not retrace at all, since he could verify by memory. While this analysis is admittedly speculative and clearly an oversimplification, it has a certain reasonableness and has the advantage of suggesting a method for the study of visual pattern perception or of problem-solving behavior in general. Specifically, it suggests that eye movements that result in the obtaining of new data represent a data-acquisition process, whereas recursive eye movements represent a forgetting process. Furthermore, it suggests that there should be a difference between those data-acquisitioning eye movements which are directed by an hypothesis-testing process and those which are made in the absence of an available concept. Finally it suggests that conditions which influence short-term memory should affect the frequency or nature of recursive movements and that eye movements, in general, should be sensitive to conditions which influence attention and, therefore, that eye-movement sequences may serve as a stress indicant. These possibilities for the use of eye-movement data as a means for studying nonvisual processes involved in ocular

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 21, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by NASA, contract No. NsG-718 with the Guggenheim Center for Aerospace Health and Safety, Harvard University School of Public Health and conducted at the Guggenheim Center; Ross A. McFarland was the principal investigator. This paper was presented to the Eastern Psychological Association, Atlantic City, New Jersey, April, 1965.

[2] Now at Tufts University.

patterns were noted by Brandt (1) in 1940 and more recently by Mackworth and Thomas (3) in 1962 and by Mackworth (2) in 1964.

Although there are available a fairly large number of eye-movement studies and quite a large number of problem-solving and concept-formation studies, there seem to be no experiments in which eye-movement data were obtained while the subject (*S*) was presented with a specific problem requiring the development of a unique concept. Our intention was to obtain a first set of data of this sort and with it to explore the usefulness of the general approach.

B. METHODS

The problems used were letter sequences reported by Simon and Kotovsky (4). Twelve of their 15 sequences were selected as those which could be classified into three groups of four each as easy, moderate, and difficult based on a preliminary experiment.

A sample of the sequences, classified as easy, is RSCDSTDETUEF—.
A sequence classified as moderately difficult is ABMCDMEFMGHM—.
A sample classified as difficult is JKQRKLRSLMST—.

The *S* was instructed that his task was to supply the next letter that would occur if the sequence were extended. Thus, in the sample sequences the correct responses are the letters, U, I, and M.

The Mackworth Eye Camera was used to obtain 10-second duration Polaroid photographs of *S*'s eye positions while he solved the problem. The sequences were back projections of 2 × 2 inch slides to a frosted glass screen normal to *S*'s view and at a viewing distance of 28 inches. The projected letters were white on a dark background. Each letter subtended a visual angle of 0.75°; the letters were 1.75° apart on center. An "X" 2° to the left of the first letter on the screen marked the position that *S* was instructed to fixate until given a verbal signal to read the sequence. The slide was made available simultaneously with this signal. The *S* was also instructed to close his eyes when he had solved the problem, thereby omitting any unnecessary eye-movement recordings. He then reported his solution verbally. The photographs contained the letter sequence, the positions of *S*'s eye during the 10-second period, and an associated time mark which permitted identifying the sequence of eye positions.

1. *Subjects*

The *S*s were 10 male college students paid for their participation.

2. *Procedures*

Each *S* was instructed first in the nature of the sequences and given a practice slide, but without photographic recording. He was then run in one session under three different experimental conditions: (*a*) unlimited time to inspect the display—unknown to *S*, photographs were taken only of the first 10 seconds of this process; (*b*) mild speed stress—*S*s were permitted only 10 seconds to view the display; (*c*) blur with unlimited time—to blur the projection the projector was slightly defocused a constant amount. The amount was that sufficient to provide readable letters, but with definite loss of contours.

To control for order effects in the use of the same sequences under different experimental conditions, the difficult and moderate problems were deliberately confounded into a counterbalanced arrangement for the unlimited and speed-stress comparisons. Half of the 10 *S*s first viewed the difficult sequences under the unlimited condition and then viewed the moderately difficult ones under speed stress; this order was reversed for the other five *S*s. The easy slides were always used for the blurred condition.

C. RESULTS

For the purpose of scoring the data, the Polaroid photographs were opaquely projected to a data form containing the sequences in actual size. The data were transferred to this from the photograph and all measurements were then made on the transferred data. A variety of measures were obtained. For each measure the data were grouped according to whether *S* solved the sequence correctly, whether the measure was in the first half or second half of the 10-second-exposure period, and whether the measure was part of a forward or recursive-movement sequence. The measures used were number of fixations, distances between fixations, and runs. A run was defined as two or more fixations in a constant direction. Thus, a run does not differentiate number of fixations in a constant direction; however, number of forward or recursive fixations does.

Comparisons were made between sequence difficulty levels, but since these had been deliberately confounded, only a small number of between-group measures were available having comparable conditions. With these there was a suggestion of an increased number of fixations as difficulty increased, but the results were not consistent.

Table 1 summarizes the results with regard to fixations, runs, and length of eye movements in terms of correctly and incorrectly solved sequences. The fixations are presented without regard to direction of movement; the lengths are in minutes of visual angle. The bottom of the table presents the mean percentage of right responses for each condition.

Across experimental conditions, the table shows that all three measures decreased systematically from unlimited to speed to blur. One sizeable inversion may be seen in the blurred-right mean for fixations, and to a smaller degree for runs. Another is in the relatively small movement length of the unlimited-wrong case. However, in general, and without exception, when right and wrong are not considered, the number of fixations, the number of runs, and the length of movements decreased systematically across experimental conditions in the order shown.

TABLE 1
MEAN EYE AIMING MEASURES

Measures	Unlimited	Speed	Blurred
No. of all fixations	16.8	15.1	14.9
Right	16.4	13.8	15.4
Wrong	17.1	16.1	14.1
No. of all runs	5.7	5.1	4.5
Right	5.6	4.4	4.9
Wrong	5.9	5.3	4.6
Length of movement (min. arc)	196	178	166
Right	200	149	148
Wrong	188	198	182
Per cent right	50	47	64

The first subdivision of Table 2 shows the number of fixations under each condition analyzed into forward and backward movements and first and second half of the 10-second recording period. The table shows that the number of forward fixations always exceeded backward ones. It can also be seen that forward fixations decreased in number from the first to the second five seconds, whereas backward movements increased in number.

The second subdivision of Table 2 shows the same breakdown in terms of runs. Although these results are not as consistent as the previous ones, and the differences not large, they also suggest a decrease in the order of unlimited, speed, and blur. They are consistent throughout in showing slightly more runs on the average in the forward direction. They also suggest differences between first and second 5-second periods, but the differences are very small.

Results in terms of length of movement are shown in the third subdivision of Table 2. The table shows that in order of unlimited, speed, and blur, the length of both kinds of movement decreased, and that forward movements decreased in length from the first to the second five seconds, whereas backward

TABLE 2
MEANS FOR THREE MEASURES DIRECTIONALLY

Exposure period	Unlimited Fwd.	Unlimited Bkd.	Speed Fwd.	Speed Bkd.	Blurred Fwd.	Blurred Bkd.
Mean number of fixations directionally						
First 5 sec.	6.5	2.3	8.1	2.4	6.0	1.9
Second 5 sec.	5.5	3.1	6.1	3.1	4.5	2.5
Mean number of runs directionally						
First 5 sec.	1.7	1.1	1.7	1.2	1.4	1.0
Second 5 sec.	1.5	1.4	1.6	1.3	1.4	1.2
Mean movement length (minutes of arc) directionally						
First 5 sec.	162	284	152	250	160	188
Second 5 sec.	156	344	140	307	125	254

ones increased. The table also shows that backward movements were always longer than forward ones.

D. DISCUSSION

All measures decreased in the order of unlimited, speed, and blur. Thus, in that order, the number of runs became smaller, the number of fixations was reduced, and the distance between fixations was reduced. This suggests that in solving problems under the kinds of mild stress that we used, *S* tends to spend less time scanning and, perhaps, hypothesis-testing and more time extracting information from each detail of what he sees. In different terms, his attentional field becomes narrowed and his mode of processing information becomes more systematic under mild stress than in nonstress conditions. Such an interpretation is consistent with data reported by Mackworth (2) showing that when portions of a picture are presented in blurred form for identification, there is a marked increase in attention to the blurred area relative to the unblurred area.

The number of fixations and the number of runs were both greater for incorrectly solved sequences than for correctly solved sequences. In general, the length of movement was longer too. Again this suggests a reduction of scanning and an increase in attention to detail associated with correct solutions.

Recursive movements were found to be considerably longer than forward ones, a finding which is consistent with the assumption that recursive movements represent instances of forgetting. Inspection of individual records did not suggest any systematic rereading or rescanning of the entire sequence as might occur if the backward movements were a systematically occurring search activity rather than a memory-refreshing one. Thus, although some of the backward movements may have represented such an activity, many or most of them can also be interpreted as representing a memory-refreshing one.

A very consistent result was the finding of more forward than backward movements in the first five seconds and more backward than forward ones in the second five seconds. If backward movements represent a memory-refreshing process, this result suggests that the initial activity of *S* is primarily data acquisition and that this is followed by a primarily memory-refreshing activity. As discussed earlier, backward movements may represent the verification of a concept or hypothesis; however, if memory were perfect, *S* would not need to look back for verification since he would remember the sequence.

E. Summary and Conclusions

The results suggest a narrowed attentional field and a heightened attention to detail with slight blurring and with mild speed stress. They suggest the same process as associated with correct solutions even in the absence of stress. The results also suggest a systematic change in the problem-solving strategy from information-gathering in the early time period to memory-refreshing and verification in the later one. While it is really too soon to be confident about the identification of these particular underlying processes as being reflected by eye-movement data, the data appear to be sufficiently systematic and sufficiently consistent to indicate that the present approach offers an important experimental method for use in further study.

References

1. Brandt, H. F. Ocular patterns and their psychological implications. *Amer. J. Psychol.*, 1940, **53**, 260-268.
2. Mackworth, N. H. Searching for visual information during the process of recognition. Paper presented to Psychonomics Society, Niagara Falls, October 1964.
3. Mackworth, N. H., & Thomas, E. L. Head-mounted eye-marker camera. *J. Opt. Soc. Amer.*, 1962, **52**, 713-716.
4. Simon, H. A., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychol. Rev.*, 1963, **70**, 534-546.

Office of Research Coordinator
Tufts University
Medford 55, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 39-40.

DOGMATISM AND TEST ANXIETY*

Columbia University

MARTIN T. REBHUN

A. INTRODUCTION

Rokeach has hypothesized that the more closed-minded a person is, the more he will have "a relatively closed cognitive organization of beliefs and disbeliefs about reality" (1, p. 195). In line with this hypothesis, we should expect that highly dogmatic people should have greater difficulty in receiving or utilizing new information than should less dogmatic people. A concomitant of this difficulty would be that highly dogmatic individuals, because of the greater personal threat involved, would show more anxiety in situations where they have to exhibit new information and knowledge—namely, in the test situation.

The present study is concerned with the relationship between dogmatism and test anxiety. It was predicted that dogmatism would be directly related to test anxiety. That is, scores on the Dogmatism Scale should correlate positively with test anxiety scores.

B. METHOD

The Rokeach Dogmatism Scale (2) and Sarason's Test Anxiety Scale (TAS) were administered to three different groups: 108 male undergraduates in a highly selective college during the fall semester, another group of 78 male undergraduates in the same college during the spring semester, and a third coeducational group of 123 graduate and three undergraduate students in a psychology course in a teachers' college.

C. RESULTS

A significant positive correlation between the Dogmatism Scale and the TAS is found in all three groups. In the fall semester undergraduate group the correlation was .25 ($p < .01$). In the spring semester undergraduate group it was .26 ($p < .05$), and in the teachers' college group it was .23 ($p < .01$).

The results are summarized in Table 1.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 27, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

TABLE 1
CORRELATIONS BETWEEN DOGMATISM SCALE AND TAS IN THREE GROUPS

Group	N	Correlation	p
Fall semester undergraduates	108	.25	$< .01$
Spring semester undergraduates	78	.26	$< .05$
Teachers' college	126	.23	$< .01$

D. DISCUSSION

The present findings demonstrate a significant positive relationship between dogmatism and test anxiety in the three groups tested. The findings support the hypothesis that closed-minded people tend to be more anxious than open-minded individuals in examination situations.

The results of the present study suggest a possible explanation for the negative relationship found between the TAS scores and measures of intelligence (3, 4, 5, 6). That is, high dogmatism may impair intellectual functioning due to anxiety generated in the learning situation—the dilemma for the highly dogmatic people being a conflict between the need to achieve and the need to exclude new information. Also, in excluding information, the performance itself is likely to be diminished. Such possibilities appear to call for further investigation into the background, early learning experience, and comparative performance of high- and low-dogmatic and test-anxious individuals.

REFERENCES

1. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61**, 194-204.
2. ————. The Open and Closed Mind. New York: Basic Books, 1960.
3. SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. Consult. Psychol.*, 1957, **21**, 485-490.
4. ————. Intellectual and personality correlates of test anxiety. *J. Abn. & Soc. Psychol.*, 1959, **59**, 272-275.
5. ————. Test anxiety and the intellectual performance of college students. *J. Educ. Psychol.*, 1961, **52**, 201-206.,
6. ————. Test anxiety and intellectual performance. *J. Abn. & Soc. Psychol.*, 1963, **66**, 73-75.

813 Hamilton Hall
Columbia College
New York, New York 10027

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 41-46.

CHILD BEHAVIOR AND PARENTAL ATTITUDES*

Southern Illinois University

MICHAEL ZUNICH

A. PROBLEM

In the development of a child's personality the influence of parental attitudes is a meaningful factor. Parental attitudes constitute the main social influence that the child experiences during his earliest years. These attitudes together with the frustrations they impose, the techniques of control, and the provocatives they offer serve as originative forces on the child's behavior, especially during the formative years. Yet in reviewing the literature one finds comparatively little empirical evidence concerned with the relation between parental attitudes and child behavior. Additional studies may aid in understanding the complexity of problems involved between parental attitudes and behavior of children. The purpose of this study was to investigate the relation between parental attitudes and the actual behavior of children.

B. METHOD

1. *Subjects*

Of the 36 subjects, 18 were female and 18 male ranging in age from 2-9 to 5-0. The parents of these subjects ranged in age from 22 to 45.

2. *Experiment*

Direct observation, utilizing a time-sampling technique and predetermined categories, was conducted in the child-development laboratory during regular school hours, through a one-way vision screen. The conditions of laboratory observations were as follows:

1. A portable, battery-operated unit with a controlled flashing light was utilized as a timing device (this apparatus was deemed very successful, since the observer could view the light and child simultaneously and could record the behavior at the exact time of the flash).
2. Each observation consisted of a five-minute duration.
3. No two observations were made on one child in succession, and the

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 3, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

number of observations was restricted to not more than two on any child during a day.

4. Observations were made during self-regulated periods and were not made in circumstances in which the teacher was conducting the activity (e.g., music or story periods).

5. An alphabetical sequence was established for observing the children. An observation was made on the next child appearing on the list if a child was absent.

The child's behavior was rated and recorded under the appropriate category every five seconds during a five-minute-observation interval. Twenty-four observations were made on each child, making a total of 960 observations of five-minute duration.

Brief definitions and descriptions of the categories used in the investigation appear below. The recording symbols, behavior categories, and definitions were:

(x) *Asking Permission*—e.g., confirms the advisability of a proposed action. Example: "May I put some of this sand in the water?"

(c) *Contact*—e.g., playing and/or conversing with others physically or verbally. Physically means sitting or being near others even though nothing is spoken. Verbally means purely social conversation with others. Example: "This is a nice doll—it is just like the one I have at home."

(+) *Cooperation*—e.g., responds to comments, suggestions, or requests with apparent interest and willingness. Example: "I would love to play house with you."

(n) *Criticism*—e.g., criticizes or blames, or becomes physically aggressive. Example: "You can't throw a ball very well, can you?" or "You're mean—I don't like you."

(d) *Directing*—e.g., specifically states the course of action he wants others to follow. Example: "Put that block on top—give me the green one—here, hold the doll while I finish the house."

(z) *Indications of Anxiety*—e.g., expresses verbally or nonverbally some concern over the consequences of his activity. Also includes emotional reaction, crying or near approach to it. Example: "I spilled some water," or "I broke the turtle."

(i) *Interference*—e.g., interferes with an activity with the intent of stopping it completely. Example: "Hey there, don't splash anymore." or "No, you are not to drink the water from those cups."

(—) *Noncooperation*—e.g., ignores, or refuses to comply with or to accept stimulation. Example: "No, I don't want to do that."

(p) *Playing Interactively*—e.g., plays with others within the framework of his own conception of play. Example: "I'll be the milkman and I'll deliver to you some milk."

(m) *Praise*—e.g., expresses approval or affection or is considerate of others' welfare. Example: "That's a nice house you built for the lobster" or "I'll wipe the water off this chair so you can sit here."

(o) *Remaining out of Contact*—e.g., plays independently away from rest of the group, or looks at books. Example: Looking out of the window.

(r) *Restricting*—e.g., attempts to modify others' behavior by qualifying its intensity, speed, and manner of executing, but does not stop activity completely. Example: "Don't splash the water so high."

(a) *Seeking Attention*—e.g., calls attention to himself or his activity; desires focusing on himself and what he is doing. Example: "Look at what I'm doing" or "Look at this road that I've built—look!"

(k) *Seeking Contact*—e.g., asks to come into physical contact with him. Example: "Come over and sit by me" or "I wish you'd come over and play in the water with me—won't you?"

(h) *Seeking Help*—e.g., physical—asks help either with some difficulty connected with himself or with his activities. Example: "I can't pound this right—hold the nail for me." Mental—asks for ideas in playing. Example: "What can I do now?" or "How can I make this bridge stay up?"

(t) *Seeking Information*—e.g., questions in pursuit of factual knowledge. Example: "What kind of boat is this?"

(m) *Seeking Praise*—e.g., indicates a desire for praise or affection. Example: "Don't you think this is a nice house that I've built?"

(s) *Suggests*—e.g., suggests an activity or politely phrases his request. Example: "Let's have a tea party now, shall we?" or "Will you bring me a towel, please?"

Prior to observations of subjects it was necessary for two observers to become familiar with the behavior categories, code letters, and scoring system. Observer reliability was based on total category appearance, and refers to the ability on the identification of an observed behavior and its placement within a given time interval.

The reliability of recording the observation was obtained by calculating the per cent of agreement between two observers recording child behavior simultaneously and independently during 30 five-minute periods. The formula utilized was number of agreements divided by 60. The measures of agreement were obtained by comparing the observations of one observer with observations taken by another observer. The number 60 represents the total number of symbols recorded in a five-minute period, one symbol being recorded at every five-second interval. Children who were not included in this study served as subjects for establishing reliability. The total observer reliabilities for the 30 periods ranged from .83 to .97. This represented a total of 150 minutes of observation.

The Parental Attitude Research Instrument (PARI) developed by Schaefer and Bell (1) was utilized to obtain attitude measures toward child rearing. The PARI was administered to the parents after the child's interaction was observed and recorded. The instrument was given separately to each parent on two different occasions. Approximately two weeks elapsed before the second parent was given the PARI.

Sixteen of the 23 subscales in the PARI were utilized. Since the PARI measures attitudes toward child rearing and family life, only subscales pertaining to child rearing were used in this study. Seven of the subscales that were concerned with family-life attitudes were excluded.

C. Results and Discussion

Examination of appearance of categorized child behavior in the laboratory showed that *Contact, Cooperation,* and *Playing Interactively* constituted the major types of child behavior while under observation. Although some of the differences were small in frequency, girls evidenced more positive types of behavior than boys, and boys more negative types than girls. The girls exhibited such behaviors as *Contact, Cooperation, Playing Interactively, Praise,* and *Seeking Information* much more frequently than boys. However boys had a higher frequency than girls of *Criticizing, Directing, Indications of Anxiety, Interference, Noncooperation, Restricting, Seeking Attention, Seeking Contact,* and *Seeking Praise.*

Significant relations which appeared between parental-attitude scores and observed-behavior ratings appear in Table 1.

Of the comparisons made between the frequencies in the 18 child-behavior categories and 16 attitude subscales, nine evidenced significant relationships at the .05 level or beyond. Relationships are suggested by the significant values of the correlation coefficients between attitude subscales and behavior categories: Subscale: *Approval of Activity;* Behavior Categories: *Contact* and *Playing Interactively.* Subscale: *Breaking the Will;* Behavior Categories: *Criticism, Directing,* and *Indications of Anxiety.* Subscale: *Comradeship and Sharing;* Behavior Category: *Contact.* Subscale: *Encouraging Verbalization;* Behavior Category: *Contact.* Subscale: *Irritability;* Behavior Category: *In-*

dications of Anxiety. Subscale: *Strictness;* Behavior Category: *Indications of Anxiety*.

In view of the 288 comparisons made between the attitude subscales and child-behavior categories these statistical relationships should be undertaken with caution as one may observe values which attained significance at the .05 level approximately five times out of 100 when no real relationship exists.

TABLE 1
CORRELATION VALUES BETWEEN PARENTAL-ATTITUDE SCORES AND OBSERVED-BEHAVIOR RATINGS

	Behavior Categories					
Attitude Subscales	Contact	Criticism	Directing	Indications of Anxiety	Seeking Attention	Playing Interactively
Approval of Activity	.63***					.33*
Breaking the Will		.34*	.42*	.42*		
Comradeship and Sharing	.58**					
Encouraging Verbalization	.68***					
Irritability				.38*		
Strictness				.33*		

* $p = .05$.
** $p = .01$.
*** $p = .001$.

Although the child-behavior categories were not specifically designed to yield measures of phenomena assessed by the PARI, many of the observed child-behavior categories were hypothesized to reflect qualities of relationships which the PARI was designed to measure. It was hypothesized, for example, that the PARI subscale *Comradeship and Sharing* would be significantly related to behavior categories *Asking Permission, Contact, Cooperation, Playing Interactively, Praise, Seeking Information,* and *Suggests*.

It also should be noted that the lack of statistically significant relationships between parental attitudes toward child rearing and most of the selected behaviors of children in a structured laboratory setting may be a function of the fact that the child's behavior may not have been representative of his behavior under home conditions. In his study of observation of mother-child interaction in the laboratory setting and in the home, Schalock (2) points out that the behavior, in many respects, is not comparable. The child, for example, may have displayed more socially acceptable behavior and may have inhibited less acceptable behavior than was usual in the home situations.

Since many of the published findings concerning parent-child relationships have relied on information obtained from questionnaires, rating scales, and interviews rather than upon direct observation, it is difficult to compare the present results. However Zunich (3) noted that there is considerable inconsistency between what parents report to be their attitudes concerning the guidance of children and their behavior when observed in interaction with their children. If the investigator obtained a reasonably close approximation of child behavior, the findings suggest that child behavior of children in a structured laboratory setting cannot be predicted from an analysis of parental attitudes toward child rearing.

D. Summary

The purpose of this study was to investigate the relation between parental attitudes and the actual behavior of children. Of the 36 subjects, 18 were female and 18 male ranging in age from 2-9 to 5-0. The parents of these subjects ranged in age from 22 to 45.

Direct observation, utilizing a time-sampling technique and predetermined categories, was conducted in the child-development laboratory during regular school hours, through a one-way vision screen. The Parental Attitude Research Instrument was administered to the parents after the child's interaction was observed and recorded.

Contact, Cooperation, and *Playing Interactively* constituted the major types of child behavior in the laboratory. Girls evidenced more positive types of behavior than boys; boys, more negative than girls.

Of the 288 comparisons made between the attitude subscales and child-behavior categories, nine evidenced significant relationships at the .05 level or beyond.

References

1. Schaefer, E. S., & Bell, R. Q. Development of a parental attitude research instrument. *Child Devel.,* 1958, **29**, 339-362.
2. Schalock, H. D. Observation of mother-child interaction in the laboratory and in the home. Unpublished Doctoral dissertation, University of Nebraska, 1956.
3. Zunich, M. Study of relationships between child rearing attitudes and maternal behavior. *J. Exper. Educ.,* 1961, **30**, 231-241.

P.O. Box 1012
Carbondale, Illinois

Published as a separate and in *The Journal of Psychology*, 1966, 62, 47-54.

PERSONALITY VARIATION IN LOWER-CLASS CHILDREN*[1]

Youth Development Center, Syracuse University

DAVID E. HUNT AND JOHN DOPYERA

A. INTRODUCTION

Riessman's description of "The culturally deprived child" (9) emphasized the differences, especially in learning style, between lower-class and middle-class children as a means of pointing out the special characteristics of the culturally deprived student to teachers and other persons working with them. His emphasis on between-class differences, however, led to a single within-class type with a single cognitive style, *the* culturally deprived child. This oversimplified representation of all lower-class children in terms of one general learning style was apparently intended to change teachers' conceptions of these children through understanding how they differed from middle-class children, but, at the same time, this monolithic view may also become a stereotype in its own right. Because of the increasing interest in planning educational programs for the culturally deprived student, and the impact of Riessman's description upon such programs, it seems appropriate to consider from an empirical standpoint the degree of difference in orientation between lower-class and middle-class children as well as the degree of variation within each class.

Therefore, the present report will consider personality variation within and between social classes from the Conceptual Systems viewpoint set forth by Harvey, Hunt, and Schroder (3) and elaborated by Hunt (4). Working on a project to develop indicators of developmental change in lower-class children, we have derived personality patterns appropriate for characterizing personality variation not only within a culturally deprived population, but also

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 5, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Research reported was supported by the Ford Foundation, by a contract with the U. S. Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program, and by research grant M-3517 from the National Institute of Mental Health, U. S. Public Health Service. Collection of data was made possible by the generous cooperation of Dr. Mario Fantini, Director of the Madison Area Educational Project; Mr. Joseph Bongo, Principal, Madison Junior High School, Syracuse, N. Y.; Mr. Ralph Westervelt, Superintendent of Oneida, N. Y. School District; and Mr. Otto Shortell, Principal, Oneida Junior High School. The assistance of Charles F. Halverson, Jay Land, and Roy E. Shore in collecting data and coding responses is gratefully acknowledged.

within other populations (5). Studying personality variation within lower-class children has the advantages of permitting improved understanding of these children and, if the dimensions can be applied to middle-class children, of opening the way for seeing similarities between lower-class and middle-class children.

From the Conceptual Systems view, development is seen as a continuous process which, under optimal conditions, proceeds in a given order to a higher conceptual level (CL). Higher CL is associated with increased interpersonal maturity and increased conceptual complexity, and thus is a desirable state. Under optimal conditions, progression on the CL dimension is viewed in terms of a series of successive stages, each characterized by a specific interpersonal orientation and conceptual structure. If environmental conditions are not optimal, then the person remains at a lower CL, and may become closed to further progression. For present purposes, we are concerned only with the person's current position on the CL dimension which can best be described through summarizing the characteristics of the developmental stages. Stage characteristics are derived from assumptions about a series of increasingly complex forms of "conceptual work" (4), and stages are assigned Roman numerals to indicate increasing CL.

Stage I work consists of defining the external boundaries and learning the generalized standards which apply to all persons. The person learns what is "right" and "wrong," he learns what is culturally expected and how to play culturally prescribed roles, and he experiences events categorically either as "good" or "bad." Stage I orientation is similar to Piaget's stage of "moral realism" in the concern with compliance to rules.

The generalized ground rules acquired in Stage I provide the foundation for the Stage II work of defining oneself and learning self-responsibility. Stage II is characterized by concern about how one is distinctively oneself and different from others, thus requiring a "breaking away" from the generalized standard. This initial expression of independence may appear in exaggerated form, but nonetheless the self-delineation produces a more highly differentiated structure than at Stage I. Stage III and Stage IV are concerned with empathic learning about others and a more complex assimilation of structure, respectively, but these higher stages occur very infrequently, if at all, in the age range of the present study so that we will be concerned only with the lower stages.

As mentioned earlier, difficulty may occur at any one of the stages. The person may not successfully accomplish Stage I work, in which case we view his orientation as a Sub I stage which is characterized by impulsivity and immaturity. When a child at Sub I stage enters adolescence, he may attempt to become more independent, without the necessary Stage I foundation which produces considerable negativism.

Rather than classify a person into one of the three stage groups as we have categorically done for other purposes, we assigned each person a single score on the continuous CL dimension. A CL score of zero indicated impulsivity, egocentricity, and inadequate socialization (Sub I); a CL score of 1 indicated compliance and dependence on rules (Stage I); a CL score of 2 was characterized by raising questions, independence, and self-assertiveness (Stage II).

Although a person's position on the CL dimension is a very general description of his personality, CL has been related to a broad domain of variables, such as differential educational treatment (5), incidence of delinquency (7), different forms of parental training conditions (1), and cognitive factors (11). Because of its relevance to both cognitive functioning and differential educational treatment, CL seemed appropriate for the present investigation; and we therefore assessed CL in a lower-class population and in a middle-class population to investigate the degree of difference and the degree of variation within each class. On the basis of common sense and earlier investigations, it was expected that CL would be generally higher in the middle-class than in the lower-class group. Some recent descriptions imply little cognitive variation among lower-class students, but there was not a sufficient basis on which to predict which class would show greater variation; so there was no explicit expectation about CL variation.

B. METHOD

1. *Subjects*

The lower-class group consisted of all students attending a "center city" junior-high school in a lower-class section of a metropolitan area, over three-fourths of whom were Negro.[2] The middle-class group consisted of all students attending a junior-high school in a nearby community which is primarily middle-class and contained very few Negro students. Data were collected for approximately 95 per cent of the students in each school, so that the lower-class group contained 297 subjects and the middle-class, 717 subjects.

Practically speaking, there was no question that these two school populations were very different in social class level even though there may have been an occasional instance of individual overlap in the two class distributions. This commonsense impression was supported by the mean grade level of father's education, which was 10.3 years for the lower-class group and 12.0 years from the middle-class group. (Because of the large number of fatherless families in the lower-class group, this information was available for only about half of this group, so that the educational level of lower-class fathers is probably considerably overestimated.)

---

[2] That this population is culturally deprived has been implicitly acknowledged in a recent decision to close this school and to assign the students to several other schools in the city in an effort to redress racial imbalance.

2. *Procedure*

Subjects were tested in classroom-size groups by two experimenters. The sentence completion method, which was administered as part of a larger battery, provided the basis for obtaining CL scores. Each subject received a booklet containing stems—e.g., "Rules," "When I am criticized"—one on each page, and was given three minutes in which to write on each topic.

Each of the six responses on the 1014 protocols was scored independently by two judges using the manual developed by Hunt and Halverson (6) specifically for measuring CL. Since this coding procedure was critical to the results, the general referents and an example for each scale point are presented in Table 1.

TABLE 1
CL SCORING REFERENTS AND EXAMPLES

CL score	Generic referents	Example from "When I am criticized"
0	Highly egocentric; no alternatives; exclusion of all negative events.	"I ignore it. I don't like it and never want to be criticized."
1	Categorical; little or no self-delineation; concern with "right-wrong," "good-bad," "should" etc.	"I should take it with thanks. I work harder and try to do the right thing."
2	Beginning self-delineation; alternatives available; greater differentiation.	"It tells me that other people may think quite different from me."
Unscorable	Clichés; insufficient referents; stem repetition.	"I am criticized."

Even though interrater reliability between trained judges is fairly high ($r = .85$ to .90), each of the 6084 responses was scored by two judges in order to maximize scoring precision. The subject's CL score was obtained by computing the average of scorable responses. If three or more responses were unscorable, the subject was classified unscorable and dropped from the analysis. Twenty lower-class and 25 middle-class subjects were dropped for this reason. Therefore, the analysis was based on 277 lower-class subjects (132 male and 145 female) and on 692 middle-class subjects (356 male and 336 female).

C. RESULTS

Mean CL scores by social class and sex group are presented in Table 2.

As expected, the lower-class mean CL scores were significantly lower than the middle-class scores for male, female, and total groups. However, in all three comparisons, the lower-class group showed significantly *more* CL variation than the middle-class group. Because of the lower CL mean score in the lower-class group, the question of whether the greater lower-class variation might have resulted from a large proportion of very low scores was considered. However, an inspection of the two class distributions indicated that, even though the lower-class CL mean was .90, there were almost as many in the lower-class group with scores above 1.0 (25 per cent) as there were in the middle-class group (30 per cent). The greater lower-class variation resulted partly from the extremely constricted middle-class variation in which almost half (47 per cent) of the subjects scored exactly 1.0—i.e., Stage I. When the data were considered in terms of stage classification, the lower-class group contained a higher proportion classified at the Sub I stage, and the middle-class group contained a higher proportion at Stage I—both as would be expected—but the lower-class group had a slightly higher proportion at Stage II.

TABLE 2
MEAN CL SCORES BY SOCIAL CLASS AND SEX

Class group	Male Mean	Male *SD*	Female Mean	Female *SD*	Total Mean	Total *SD*
Lower	.79	.36	.99	.33	.90	.36
Middle	1.01	.28	1.06	.22	1.03	.25
t	7.10**		2.77**		6.43**	
F		1.30*		1.50**		1.42**

* Significant beyond the .05 level.
** Significant beyond the .01 level.

Sex differences were also analyzed; and in agreement with earlier investigations on similar cognitive factors (e.g., 2), the female mean CL score was significantly higher than the male score ($p < .01$) for both class groups. In the middle-class group, the variation within the male group was significantly greater ($F = 1.26$; $df = 355, 335$; $p < .05$) than the variation within the female middle-class group, but there was no sex difference in variability within the lower-class group.

D. DISCUSSION

The relatively higher conceptual level for the middle-class group in general is in keeping with earlier results (2) and also indicates the general differences between lower-class and middle-class students as implied by advocates of special programs for the culturally deprived student. However, the finding of greater variation in the lower-class group raises questions about using a single educational prescription for the culturally deprived student.

Based on a description of the culturally deprived student as quite concrete and externally oriented, Riessman has suggested in a recent conference that the optimal educational environment for the culturally deprived student should consist of the following: "Consistency, structure, and order are fundamental. . . . Extrinsic rewards and punishment are understood." (10, p. 98). Riessman's description is similar to the Sub I orientation, and we have found that an educational environment similar to the one he describes is indeed most effective with this group (4). However, as Miller suggested at the same conference from which the Riessman suggestion came, "People vary. From what I have been able to learn, it appears that certain methods work very well with certain groups, but they do not work very effectively with others" (8, p. 119). Put in present terms, although Riessman's suggestion may be very effective for some culturally deprived students, it may be very inappropriate for others.

Using a Conceptual Systems change model (4), we have obtained evidence for the differential effectiveness of educational environments according to the student's conceptual level. With the present population of lower-class students, classroom groups were formed homogeneous in conceptual level—a Sub I group, a Stage I group, and a Stage II group. In an exploratory study with ninth-grade students, the teachers were unaware of the nature of the groupings, but they quickly discovered that the groups were differentially receptive to specific educational procedures. The Sub I group required considerable structure and concrete experiences, and could not maintain a discussion. The Stage I group functioned most effectively when competition and debates were involved. The Stage II group seemed to function most effectively when there was considerable opportunity for independent exploration, questioning, and differences of opinion. In the later study, homogeneous classroom groups were formed in all three grades, and served as the focal point in a program of differential educational treatment for the culturally deprived population. Within this context, Riessman's recommendations might be very appropriate for a Sub I group, but very inappropriate for a Stage II group (this problem is further complicated because some teachers find it difficult to distinguish between the Sub I and Stage II orientation, perhaps because both involve some opposition).

Therefore, it seems important to qualify Riessman's description and suggestion in light of the considerable variation observed within this population. If it is agreed that an understanding of a student's orientation permits deriving optimally effective educational procedures, then the present finding of considerable variation within the lower-class group is important because it points to the need for variation in the educational environments developed, and the coordination of these environments with the students. Such coordination or "matching" of student and educational environment is also applicable to middle-class students, but perhaps its most interesting application might be to a recently desegregated school where variation in conceptual level might be used as a means of planning optimal educational environments in schools containing both lower- and middle-class students.

E. SUMMARY

Riessman (9) has described the culturally deprived child in terms of a fairly uniform cognitive style, and suggested educational procedures appropriate for such students. Within this context the present investigation of junior-high school students found, as expected, that the conceptual level (CL) of a lower-class group was lower than that of a middle-class group. In contrast to what might be expected from Riessman's description, however, the lower-class group showed *more* CL variation than the middle-class group. This greater lower-class variation was discussed in terms of providing differentially effective educational procedures for students varying in CL, and the utility of Conceptual Systems theory for providing such coordination was considered.

REFERENCES

1. CROSS, H. J. The relation of parental training conditions to conceptual level in adolescent boys. Unpublished Doctoral dissertation, Syracuse University, Syracuse, New York, 1965.
2. GOLLIN, E. S. Organizational characteristics of social judgment: A developmental investigation. *J. Personal.*, 1958, **26**, 139-154.
3. HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. Conceptual Systems and Personality Organization. New York: Wiley, 1961.
4. HUNT, D. E. A conceptual systems change model and its application to education. In O. J. Harvey (Ed.), *Cognitive Flexibility*. New York: Springer, in press.
5. HUNT, D. E., & DOPYERA, J. Indicators of Developmental Change in Lower-Class Children: Interim Progress Reports. Syracuse, N. Y.: Syracuse Univ. Youth Devel. Center, 1963-1964.
6. HUNT, D. E., & HALVERSON, C. F. Manual for scoring sentence completion responses for adolescents. Unpublished manuscript, Syracuse University, Syracuse, New York, 1964.
7. HUNT, D. E., & HARDT, R. H. Developmental stage, delinquency, and differential treatment. *J. Res. Crime & Delinq.*, 1965, **2**, 20-31.
8. MILLER, S. M. A search for an educational revolution. In C. W. Hunnicutt (Ed.),

Urban Education and Cultural Deprivation. Syracuse, N. Y.: Syracuse Univ. Press, 1964. Pp. 113-126.

9. RIESSMAN, F. The Culturally Deprived Child. New York: Harper, 1962.
10. ———. Teachers of the poor: A five point plan. In C. W. Hunnicutt (Ed.), *Urban Education and Cultural Deprivation.* Syracuse, N. Y.: Syracuse Univ. Press, 1964. Pp. 95-111.
11. WOLFE, R. The role of conceptual systems in cognitive functioning and varying levels of age and intelligence. *J. Personal.,* 1963, **31**, 108-123.

Youth Development Center
Syracuse University
Syracuse, New York 13210

Published as a separate and in *The Journal of Psychology,* 1966, **62**, 55-66.

THE EFFECTS OF EXPERIMENTALLY INDUCED INITIAL SUCCESS FOLLOWED BY FAILURE OR THE REVERSE ON SELF-CONTROLLED DELAY BEHAVIOR IN RETARDED AND NORMAL SUBJECTS*[1]

Behavioral Disabilities Center, University of Wisconsin

JOHN PFLAUM

A. INTRODUCTION

The traditional delayed-reaction experiment involves either children, adults, or animals that are shown as many as five containers, one of which is baited with candy or fruit. After being shown the containers, the subject is diverted for a fixed length of time and then is tested to see if he will choose the correct container. A simplified example would be to bury or hide an object in full view of an animal or child and then see how long the subject can remember where the object was placed.

One of the first experiments on this topic was conducted by W. S. Hunter (7). Hunter tested the ability of both rats and children to delay responses. Rats were first conditioned to run to one of three lighted doors in order to get food. Then, after the light had been flashed, the rat was held back, and it was observed how long the rat could be delayed and still choose the correct door. Maximum delay was reported to be 10 seconds.

Other researchers have studied delayed reactions. The research of Harlow (5), Tinkelpaugh (11, 12), Yerkes (13), and Kohler (8) tends to indicate, according to Osgood (9), that length of delay interval increases with the phylogenetic level of the organism. It seems that the more highly-developed organisms have longer short-term memories.

But, according to M. R. Denny (4), within the human species no consistent relationship has been demonstrated between maximum length of delay and mental age, chronological age, or intelligence quotient. Barnett, Ellis, and Pryer (1) report that defectives will delay longer if they first learn names

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 4, 1965, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Funds to pay subjects were provided by the School of Education, University of Wisconsin. Computer time was donated by the research committee of the University of Wisconsin Computing Center. Mrs. Sinclair, Mr. Allen, and Dr. Stockton and Dr. Blessing were very helpful in arranging for the availability of subjects.

for the cues used in the task. Baumeister and Ellis (2) report that when exposed to a distraction, retarded subjects improve their performance and respond more quickly than retardates matched on *MA*. Most of the research on retarded subjects has been conducted with a restricted mental-age range, usually within an institutionalized population.

The experiment herein undertaken follows procedures similar to the classical studies of delayed response. A multiple-choice situation is presented to the subject who is required to choose, among four alternatives, a figure identical to one that has previously been presented. The classical procedures have been modified in order to permit the subject to control his own delay period.

The longer the subject delays, the more he is rewarded, provided the correct alternative is selected. The experimenter manipulates correct and incorrect choices by arbitrarily passing the subject on five easy items, after which the experimenter fails the subject on five difficult items, or the reverse. The purpose of these modifications of the classical delay experiment is to study the changes in delay behavior as a consequence of initial success or failure and the combination of success and failure experiences.

According to social learning theory (10), it has been hypothesized that retarded persons have a higher general expectancy for failure than do normals (3). If this were the case, given X amount of initial general expectancy for failure, each subsequent success experience would subtract Y amount of expectancy for failure. Each successive success experience would reduce the subject's specific expectancy for failure on the next item.

When failures follow successes or successes follow failures, new conditions are introduced, namely the expectation for successes following failures or the reverse. The studies of Gardner (1958) and Ringelheim (1958) as discussed by Cromwell (3) indicate that retardates do less well than normals during a failure sequence which follows a success sequence. What Gardner calls partial failure—that is, failure following initial success—is reported to be more debilitating than complete failure. The concept of expectancy-outcome discrepancy is used to explain the debilitating effect of partial failure [Gardner (1957) in Cromwell (3)]. According to the expectancy-outcome discrepancy theory, retardates (depending upon the degree of retardation) expect to fail before they succeed. Normal subjects would be more likely to expect initial success than would retarded subjects.

Social learning experiences would tend to dispose the retardate to expect initial failure. The normal is more likely to have experienced both initial success and initial failure. Since retardates have had fewer experiences of initial success than have normals, the former would be less capable of reacting appropriately in such a situation.

The appropriate behavior in this experimental situation is to increase delay time under success conditions and to reduce delay time under failure conditions. Since subjects are rewarded for increasing delay under success conditions, delay time would be expected to increase, while under failure conditions delay is not rewarded and would be expected to decrease.

While retarded subjects may be able to respond appropriately when initial failures are followed by successes, which are experiences often encountered in their daily lives, the normal subjects are more likely than retardates to have experienced initial success and hence would be expected to behave appropriately when either successes are followed by failures or the reverse (see Figures 1 and 2).

Subjects are rewarded one cent for every second delayed on the relatively easier passed items and not given any reward on the relatively more difficult

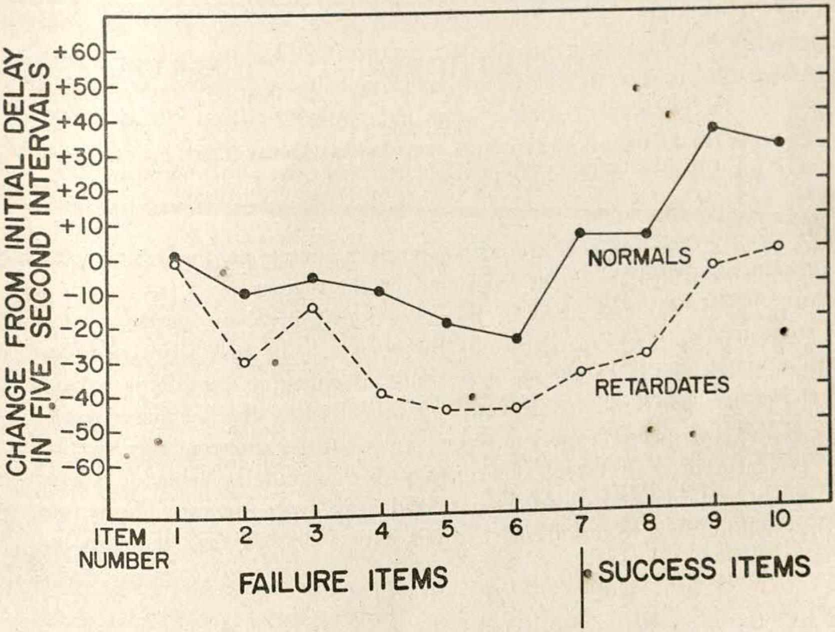

FIGURE 1
CHANGES (IN FIVE-SECOND INTERVALS) FROM INITIAL DELAY TIME, ON FIRST TRIAL, FOR SIX FAILURE AND FOUR SUCCESS ITEMS WITH 10 MENTALLY RETARDED ADOLESCENTS AND 10 NORMAL CHILDREN

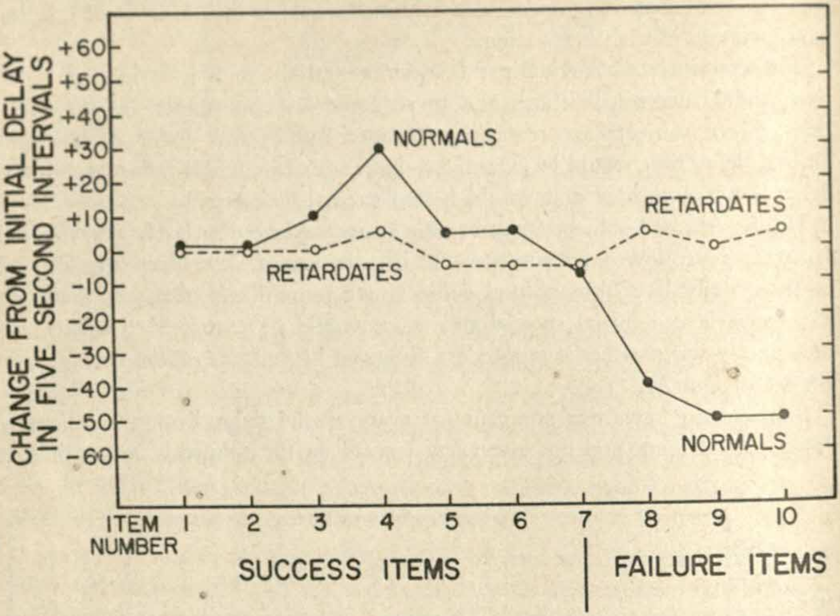

FIGURE 2
CHANGES (IN FIVE-SECOND INTERVALS) FROM INITIAL DELAY TIME, ON FIRST TRIAL, FOR SIX SUCCESS AND FOUR FAILURE ITEMS WITH 10 MENTALLY RETARDED ADOLESCENTS AND 10 NORMAL CHILDREN

failed items. The subjects are purposely given success on the initial five items and failed on the subsequent five items, or the reverse.

The particular task used in this experiment involves success experiences followed by failure experiences, or the reverse. The task is intended to test the subject's ability to change as a result of changing conditions. Change in delay behavior would be expected to occur when the behavior is at first rewarded by success experiences and subsequently nonrewarded by failure experiences, or the reverse. The capacity to change delay behavior is hypothesized to be a function of previous social learning experiences. The purpose of the experiment is to test this hypothesis with a defined task and sample.

B. SUBJECTS

Twenty retarded and 20 normal subjects were selected from the summer classes of the Madison public schools. The 20 retarded subjects were enrolled in a voluntary summer-school program for educable mentally retarded special-class pupils. All the mentally retarded subjects had been assigned to special classes during the regular school year. They had all been individually tested with the WISC and all but one with the Stanford-Binet. The mean full-scale WISC *IQ* was 70 with an *SD* of 8.2 and a range of 59 to 87. The mean Binet *IQ* was 73 with an *SD* of 9.7 and a range of 53 to 86.

The normal subjects were all taken from the summer session of the Madison public schools. Intelligence-test data were not available on over half of the normal subjects; however, it was the policy that only children of normal or above normal ability be admitted to the summer program.

The retarded subjects averaged 14-6 years old (*SD* of 1.9 months and range of 11 to 17 years old) while the normal subjects were on the average 9-3 years old (*SD* 3.9 months and range of 8 to 9 years old).

Adolescent retarded subjects were used in order to insure that instructions were comprehended and that subjects would be capable of performing the task. It was expected that the groups would be similar in their capacity to do the task but would differ in the manner in which the task would be performed. Although the two groups were expected to be comparable in length of delay and in accuracy in identifying the correct alternative, it was predicted that they would differ in their change in length of delay time when successes followed failures or the reverse.

C. Apparatus and Procedure

The materials consisted of five-inch by three-inch white index cards which were lined on one side and blank on the other side. Geometrical figures were cut out of the index cards. The figures were four-column bar graphs. The height of the bars was determined by reference to a table of random numbers.

In the multiple-choice identification task, the subject was exposed to one of these bar graphs and then, after a period of delay, four alternatives were presented, one of which was identical to the original. For the relatively easy items, on which the subject was deliberately passed, two bars of the alternatives were altered, one-half inch each, while on relatively difficult items half as much was added to or subtracted from the height of the bars of the alternative figures. Difficult alternatives were constructed by adding or subtracting one-fourth inch from two columns, while relatively easy alternatives were constructed by adding or subtracting one-half inch from two columns.

These white bar-graph figures were placed in a loose-leaf notebook with a transparent cover backed by black construction paper.

The subject was shown one white bar graph centered on a black background for five seconds. The next page, a black piece of paper, was flipped over to cover the figure; the subject waited as long as he wished up to 30 seconds. When the subject requested, the alternatives were flipped into view. The four alternatives were symmetrically placed on a black background. One alternative was exactly like the original, while the other three were different (see Figure 3).

The subject selected the alternative he believed to be most like the original. The experimenter then rewarded or did not reward the subject and went on to the next item. These procedures were developed in a preliminary pilot study.

The apparatus consisted of stimulus items, alternatives, and black dividers, all bound in a loose-leaf notebook. There were two notebooks, one in which relatively easy items came first and relatively difficult items came last, the other with the reverse order. Three example items preceded the testing in both cases. Ten test items were presented to each subject.

When the five difficult items were presented first and the five easy items presented last, the subject was arbitrarily failed on the difficult items and was arbitrarily given success on the easy items. When the five easy items were presented first followed by the five difficult items, the subject was automatically given success on the five easy items and automatically failed on the difficult items.

Half the subjects, both normal and retarded, experienced five failure items followed by five success items while the other half experienced five success items followed by five failure items. The experiment involved a 2 × 2 factorial design with an equal number of subjects in each of four cells.

While success and failure were arbitrarily manipulated by the experimenter, the responses of the subjects were recorded, and it was therefore possible to assign a score to each subject for his performance in terms of correct and incorrect choices of alternatives. The total number of figures correctly identified was determined for each subject (see Table 1).

The subject was rewarded for delay on the success items but not rewarded for delay on the failure items. The subject earned one cent for every second delayed on the success items, while no reward was given on the failure items. In other words, the subject was paid for waiting immediately after those items on which he was passed and not paid for those items on which he was failed.

The period of self-controlled delay was timed by the experimenter with a stopwatch. The time period started when the experimenter covered the original stimulus and ended when the subject signaled that he wished to see the alter-

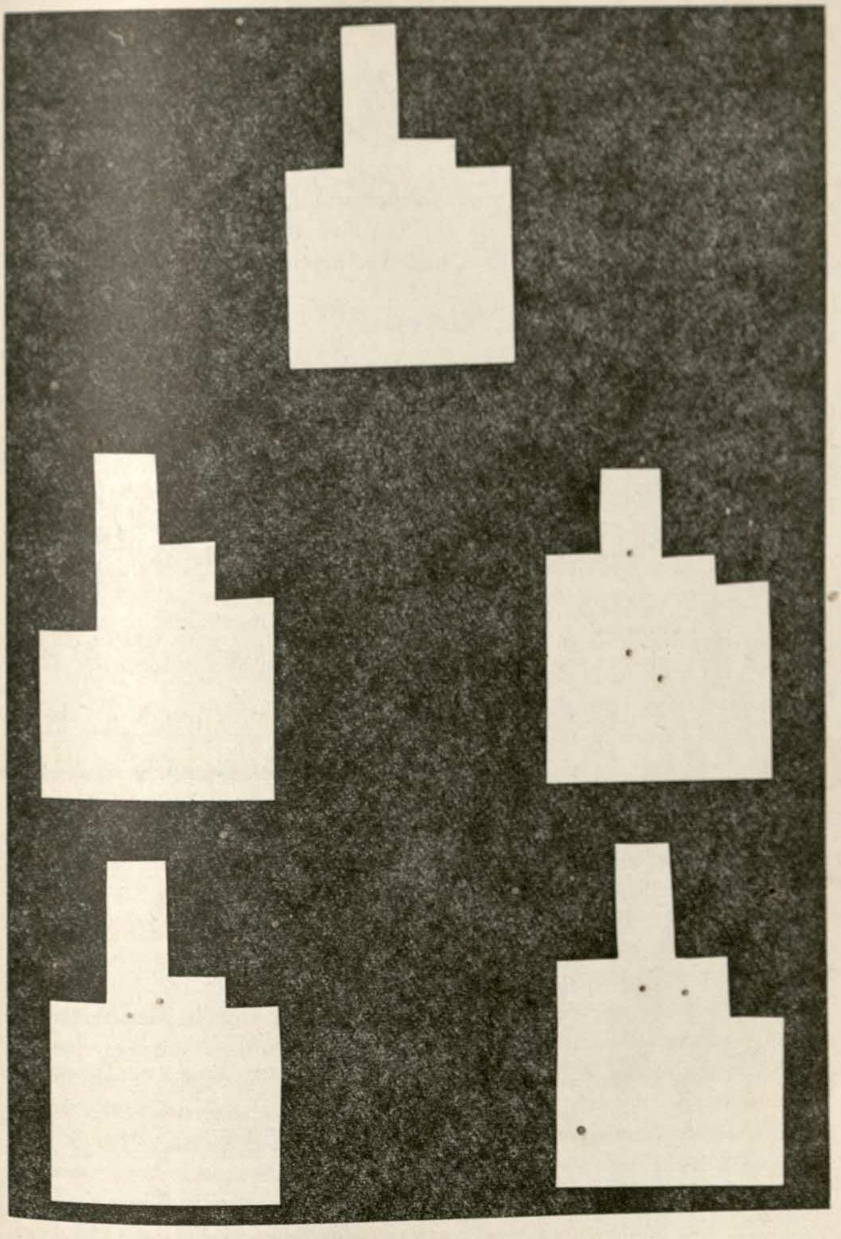

FIGURE 3
AN EXAMPLE (NOT A TEST ITEM) OF THE ORIGINAL STIMULUS AND FOUR ALTERNATIVE FIGURES, ONE OF WHICH IS CHOSEN BY THE SUBJECT AS BEING MOST LIKE THE ORIGINAL
(The original stimulus is shown here as the uppermost figure)

natives or when the 30 seconds had passed. The subject would say "now" when he wanted to see the multiple-choice alternatives. When he selected the alternative which he believed to be the same as the original, his choice was recorded, and the experimenter would say "right" or "wrong" and go on to the next item. The experimenter would either say "right" to the first five items and "wrong" to the last five items, or the reverse. The "right" items were the easy items, and the "wrong" items were the difficult items.

D. Results

The average number of correct responses for the 20 retarded subjects was 3.7. The normal subjects averaged 4.0 correct responses. The difference between the normals and the retardates was not significant ($t = .28$). There was no significant difference between the retardates and the normals in their ability to identify the correct figure. Both difficult and easy figures were identified equally well by retarded and normal subjects. The two groups were comparable in their ability to perform correctly on the task itself (see Table 1).

The difficult items were statistically significantly more difficult than were the easy items for both the retarded and the normal groups ($t = 7.6$) above the one per cent level of confidence (see Table 1).

Retardates and normals were also comparable in their length of delay on either initial success or failure conditions (see Table 2).

It was expected that normals and retardates would increase delay time under success conditions which followed failures. Ten normal subjects and five retarded subjects changed in the expected direction. When the amount of change in seconds is accounted for, the difference is not significant above the five per cent level. Change behavior over items is described in Figure 1. Since the switch from success to failure or the reverse occurs at the end of item 6, item 6 itself must be included under the previous conditions.

It was also expected that normals would decrease delay time under failure

TABLE 1
Accuracy of Identification of Five Difficult Versus Five Easy Figures for 20 Retarded Adolescents and 20 Normal Children

Subjects	Mean number correct on 5 easy items	Mean number correct on 5 difficult items	Mean number correct on total of 10 items
Retardates ($N = 20$)	2.8	.9	3.7
Normals ($N = 20$)	2.8	1.2	4.0
Total group ($N = 40$)	2.8	1.1	3.9

TABLE 2
DELAY TIME IN SECONDS UNDER SUCCESS AND FAILURE CONDITIONS FOR 20 MENTALLY RETARDED ADOLESCENTS AND 20 NORMAL CHILDREN

Subjects	Mean time of delay under success conditions	Mean time of delay under failure conditions	Mean total time of delay
Adolescent retardates ($N = 20$)	41.5	38.4	79.9
Normal children ($N = 20$)	52.4	37.6	90.0
Total group ($N = 40$)	46.9	38.0	84.9

conditions that followed successes while retarded subjects would not do so as often and to as great an extent. Out of 10 normals, all 10 changed in the expected direction while only three of the 10 retardates changed in the expected direction. The difference is statistically significant above the one per cent level ($X^2 = 9.87$, $df = 1$). Change behavior over items is described in Figure 2.

E. DISCUSSION AND CONCLUSIONS

The results tend to substantiate the expectations as derived from social learning theory, at least for the particular task and the particular subjects used. Since the sample size is rather modest and the task rather specific, any broad generalizations would be unjustified. It can be said, however, that the retarded and the normal subjects—in this sample and on this task—demonstrated the expected changes in delay behavior.

The traditional delay-reaction experiment (7) was modified so as to compare retarded and normal subjects on their capability to delay responses under initial success or failure conditions.

It was hypothesized that retarded individuals usually experience failure before success, while normals experience initial success relatively more frequently than do retardates. The situation wherein failure comes first and success follows is more realistic in terms of the life experiences of both the normals and the retardates. A situation of initial success followed by failure is less likely to have occurred in the life experiences of the retardate than in past life experiences of the normal.

If an individual is rewarded in proportion to the length of the self-controlled delay under success conditions and not rewarded under failure conditions, it can be expected that length of delay will increase under success conditions and decrease under failure conditions.

This expected adjustment to changing conditions will occur, however, only if through similar social experiences the subject has learned this form of adjustment.

When initial failure is followed by success, both retardates and normals can adjust as expected because they both have had similar experiences in previous social learning. But when initial success is followed by failure, adequate adjustments cannot be made by retarded subjects because previous social learning is not consistent with such an experience.

F. Suggestions for Future Research

It would be interesting to compare a group of normal subjects to a group of retarded subjects matched on chronological age. It would be predicted that there would be a significant difference between the retardates and the normals on the number of correct identifications. Normals would be predicted to do better on the task than would retarded subjects of equal chronological age. Since no consistent relationship has yet been found between delay behavior and mental age, this experiment would contribute to a definition of the relationship if such a relationship does exist.

A cross-sectional study of delay behavior would be quite fruitful. Such a study would establish the relationship, if any, between chronological age and capacity to delay.

The entire task could be mechanized. A tape recorder could give instructions; the subject could choose alternatives by pushing one of four buttons. The figures could be presented on slides. The subject could be shown the alternatives when he pulled a switch. Payment could be mechanized by using a transparent box into which a penny would drop for each second the subject delayed but would open only if the correct alternative were selected. Time records could be kept automatically by use of specially wired time clocks.

The task could be modified so as to have more success and failure items or to have more five-item experiences. Payment could be made in poker chips (a symbolic reward) which could be cashed in for a tangible reward at the end of the session. Different kinds and amounts of rewards could be compared for their relative influence on delay behavior. The subject could choose his own reward by ranking objects according to preference. High and low preference rewards could be compared (6).

Reward schedules could be manipulated such that the subject would be rewarded in a certain fixed or variable ratio. It would also be possible to give the subject a certain amount of money initially and then take away pennies when he fails and add pennies when he succeeds. Either manipulating the success and failure experiences or playing the game without interference, allowing natural successes and failures to occur, could be attempted.

G. SUMMARY

The experiment contrasted retarded and normal subjects in terms of their change in delay behavior when initial successes were followed by failure, or the reverse.

Twenty normal children were compared with 20 adolescent retarded subjects. Three examples plus 10 items were presented to each of 40 subjects. After the subject was shown a geometrical figure for five seconds, the figure was removed. Then, after the subject had waited as long as he wished (up to 30 seconds), four alternatives were presented. The subject's task was to choose the alternative figure most like the original. After each item, one penny was earned for each second delayed between original exposure and selection of alternatives. A maximum of 30 cents could be earned on each success item. In one treatment condition, five success items were followed by five failure items. In the other treatment condition, five failure items were followed by five success items. The subject was arbitrarily given success or failure, although his actual responses were recorded and scored.

It was expected that normal and retarded subjects would tend to delay longer on success items that followed failure than on the previously presented failure items. The normals, but not the retardates, were expected to delay less on failure items than on the previously presented success items. The results tend to confirm the predictions. The retarded subjects in this sample displayed the expected lack of change behavior. A theoretical explanation of this relative lack of change behavior is attempted within the framework of social learning theory. Social learning, from everyday life occurrences, would predispose the retardates to expect success only after failure experiences, while normal persons are more likely than retardates to have had initial success experiences.

REFERENCES

1. BARNETT, C. D., ELLIS, N. R., & PRYER, M. W. Stimulus pretraining and delayed reaction in defectives. *Amer. J. Ment. Defic.*, 1959, **63**, 104-111.
2. BAUMEISTER, A. A., & ELLIS, N. R. Delayed response performance of retardates. *Amer. J. Ment. Defic.*, 1963, **67** (No. 5), 714-722.
3. CROMWELL, R. A social learning approach to mental retardation. In N. R. Ellis (Ed.), *Handbook of Mental Deficiency*. New York: McGraw-Hill, 1963. Pp. 41-91.
4. DENNY, M. R. Research in learning and performance. In H. A. Stevens & R. Heber (Eds.), *Mental Retardation*. Chicago: Univ. Chicago Press, 1964. Pp. 100-142.

5. HARLOW, H. E., & ISRAEL, R. H. Comparative behavior of primates: IV. Delayed reaction in subnormal humans. *J. Comp. Psychol.,* 1932, **14**, 253-263.
6. HEBER, R. Personality. In H. A. Stevens & R. Heber (Eds.), *Mental Retardation.* Chicago: Univ. Chicago Press, 1964. Pp. 143-174.
7. HUNTER, W. S. Delayed reaction in animals and children. *Behav. Monog.,* 1913, **2**, 1-86.
8. KOHLER, W. The Mentality of Apes. New York: Harcourt, Brace, 1925.
9. OSGOOD, C. E. Method and Theory in Experimental Psychology. New York: Oxford, 1953.
10. ROTTER, J. B. Social Learning and Clinical Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1954.
11. TINKELPAUGH, O. L. An experimental study of representative factors in monkeys. *J. Comp. Psychol.,* 1928, **8**, 197-236.
12. ———. Multiple delayed reaction with chimpanzees and monkeys. *J. Comp. Psychol.,* 1932, **13**, 197-236.
13. YERKES, R. M., & YERKES, D. N. Concerning memory in the chimpanzee. *J. Comp. Psychol.,* 1928, **8**, 237-271.

P. O. Box 1215
Madison, Wisconsin

Published as a separate and in *The Journal of Psychology*, 1966, 62, 67-82.

AN EXPERIMENTALLY GENERATED SET OF EXPRESSIVE PHOTOGRAPHS*[1]

Veterans Administration Hospital and University of Kentucky, Lexington, Kentucky

JOSEPH LYONS AND STANLEY GOLDMAN

I. INTRODUCTION

A. PROBLEMS

A number of persistent problems—technical and practical as well as theoretical and methodological—are evident in the extended history of the study of social perception.[2] Among these are 1) generating and recording the expressions to be perceived (stimulus materials) and 2) defining the stimuli.

1. *Generating the Stimuli*

Expression is a phenomenon that is unique in many ways, and therefore its experimental production poses special problems. It is desirable that expressions used as stimuli be spontaneous as well as reproducible for experimental purposes, although these two requirements may conflict. Since spontaneous expressions are usually fleeting and unrepeatable, special methods are called for in order to capture them for objective, repeated presentation. The earliest studies utilized drawings by an artist (1, 5, 19), and when photographic techniques were introduced, they were at first combined with drawings (21).

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 7, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The authors' thanks are due to many persons—to Prof. Joseph Nuttin and Prof. L. Knops for assistance in the Louvain studies; to George Thinès, Hans Orlemanns, and Mlle. M. Nahmen for translating instructions and protocols; to David L. Horton for arranging for *S*s; to Selwyn Zerof and William Allen for computer analysis; to James S. Calvin for advice on experimental design and treatment of data; to Beverly Rose, Linda Welch, Betty F. Meketon, and Sandra Goode for assistance in organizing the data and preparing the manuscript; to F. D. Wallace and William Logan, of the hospital's Medical Illustration Service, for extensive photographic services; and to Erwin W. Straus for fruitful concepts and hypotheses throughout.

[2] The term *social perception* first appeared in *Psychological Abstracts* in 1945, although the history of interest in the problem goes back to Aristotle. The term is used here to refer to the perception of animate targets, such as faces, as opposed to the perception of inanimate targets, such as lines, colors, or shapes. Most studies of social perception have been concerned with the expression of emotion, perhaps due to the influence of contributions by Bell (1) and Darwin (5). In addition, there are significant areas of behavior usually termed "nonemotional" whose expressive characteristics have not been explored; the present study deals with such an area.

More recently, motion picture photography (4, 15, 17, 22) combined with stroboscopic lighting (24) has made it possible to capture brief and transient expressive behavior while minimizing the influence of the recording technique on the actor. However, the laboratory production of expression, while solving many technical problems, usually involves some measure of control over the actor, so that what is gained in reproducibility may be lost in "genuineness" and spontaneity (23). Efforts to deal with this problem have been made by seeking out candid records of spontaneously occurring situations (12, 18), but here again serious practical difficulties must be overcome.

2. *Defining the Stimuli*

In any experiment it is necessary that the stimulus be explicitly defined by *E* prior to presenting it to *S*. Otherwise, as Bruner notes (3, p. 131), the *E* who must depend on *S* for a definition of the stimulus is arguing in a circle. Attempts to resolve this issue in studies of social perception have taken a variety of forms: (*a*) composing the stimulus by some more or less artificial method, either by drawings (21), by breaking down the expression into units and reconstructing it as a kind of jigsaw puzzle (2, 11, 19), or by studying the effect of artificially transposing parts of the face (4, 6, 12, 20); (*b*) arranging that the behavioral situation calls forth a specific expression, so that the definition of the stimulus becomes a function of the definition of the precipitating situation (4, 7, 12, 14, 17, 18, 22); (*c*) utilizing the services of an actor to pose for the required expression, thus basing the definition of the stimulus on the actor's intent (8, 9, 16, 20, 21); (*d*) utilizing the consensus of one group of judges as a basis for preparing, selecting, or evaluating stimuli to present to other judges (10); and (*e*) selecting stimuli out of a wide range of materials on the basis of *E*'s best judgment (9, 16, 20, 24).

These approaches are not exclusive and, in fact, many studies involve various combinations of methods. We propose that basic to a workable solution to the problem of defining the stimulus must be a recognition that it is essentially a multi-stage, "boot-strap" operation. Ideally, *E* begins by selecting materials that are technically as adequate as possible, and then deciding which of the stimuli (photographs) to use; providing for consensus among independent judges in order to establish reliably that the stimuli are discriminable; utilizing consistency and consensus of judgment by additional *S*s to select stimuli for further study; and presenting these stimuli, in turn, in a sequence of studies with the dual aims of refining the stimuli and specifying the operations involved in *S*s' judgments. It is only at the conclusion of an extended sequence of such studies that the concept of accuracy of social perception can have any experimentally based meaning.

Central to this multi-stage sequence is the concept of reliability of the stimulus material. Two measures of reliability are available, and probably necessary, in the course of refining the stimuli for experimental use: (*a*) agreement (consensus) among judges and (*b*) consistency of judgment as assessed by a test-retest procedure. Both measures were utilized in the present study to assess judges' responses.[3] Together, the two methods form the basis for the screening procedure which is presented and discussed below.

B. Outline of the Study

The present study consists of the preparation of special stimulus materials, a pilot study to test out the materials, and a sequence of experiments in which stimuli were assessed and selected in terms of reliability.

1. *The Stimulus Materials*

Ten male and 10 female hospital employees were seen individually and their behavior photographed on negative stock, by means of a 16 mm motion picture camera, against the neutral background of a light-colored screen. In accordance with the standard procedure in our laboratory, the camera was in full view of the behaver at all times. Five male and five female behavers were required to perform a series of tasks involving serious, directed thinking. The other 10 behavers were given a series of situations so chosen as to elicit selected characteristics of the expression of thinking but in nonthinking situations.

2. *"Thinking" and "Nonthinking" Tasks*

A sequence of "thinking" tasks was presented in the same order for each of 10 behaver *S*s, with the explanation that *E* was developing a new series of intelligence tests for adults of superior ability. A total of 13 problems requiring serious, directed thinking was used—for example, to solve mentally a problem in long division or to manipulate in imagination a design made up of letters in various orientations.[4]

The 15 tasks presented in the same order to each of the 10 "nonthinking" behavers utilized the same neutral background as in the "thinking" tasks. The explanation offered to the nonthinking behaver *S*s was that *E* was developing a series of tests of the various sensory modalities. Tasks were chosen so that the accompanying facial expression would approximate some characteristic typically found in the expression of serious, directed thinking. For example, as an approximation of the "faraway" look often seen in the eyes of a person engaged in serious thinking, *S*s were asked to look out a window at a wall 100 feet distant and to count very rapidly the number of bricks alongside a specific window. It should be noted that pairing "thinking" and "nonthinking" expressions which were obtained in this way resulted in a discriminative task of some subtlety for the judges in subsequent experiments.

---

[3] The familiar concept of *stereotype,* which implies both intersubject agreement and some consistency over time, is, of course, relevant here.

[4] The materials used in these tasks are from unpublished tests of creative ability which were developed by Joseph Lyons and Erwin W. Straus.

3. *Sets of Pictures*

For each of the "thinking" behavers, 25 individual frames were selected from the available footage as the "best" examples of thinking and made into 2 × 2 slides. This procedure inevitably involved the selective judgment of the experimenter (J. L.), but the choice was by no means difficult to make; far more material was available than could be used. The film of the "nonthinking" behavers was similarly reviewed and, for each behaver, 15 single frames were selected as representing the "best" examples.

The 400 pictures thus selected were now used to form six sets of photographs, each set of 20 consisting of one picture of each of the 10 thinkers and one picture of each of the 10 nonthinkers. Four of these sets contrasted a characteristic of the "thinking" expression, such as vagueness, with a related characteristic found in "nonthinking" expressions; one set depicted gestures customarily seen in thinking, such as pursed lips; and one set paired thinking expressions with deliberate imitations of them.

4. *Pilot Study*

In a pilot study carried out at the University of Louvain, Belgium, French-speaking and Dutch-speaking judges were required to choose which of two photographs of a pair was "engaged in serious thinking."[5] Striking differences were indicated between the sets of pictures, the response measure consisting of judges' agreement with *E* in classifying a picture as "thinking." No differences were apparent between male and female judges (17).

On the basis of consensus among the initial 75 judges, an additional set of 20 pictures was selected for presentation to the remaining 103 judges. Ten of the pictures were those chosen most frequently as thinkers, and 10 were those chosen least frequently; five of each 10 pictures had originally been categorized as "thinker" and five as "nonthinker." The Pearson r between the first and second "score" for each picture (the percentage of times the picture was chosen as a "thinker"), when judged by 103 new *Ss* (70 male and 33 female), was $+.81$. In no instance was a picture changed from a high-choice to a low-choice category, or *vice versa,* when judged by the new sample. By this method, then, the reliability of a selection of the stimulus materials was initially demonstrated.

---

[5] The exact phrase used with the French-speaking subjects was "en train de réfléchir sérieusement," and for the Dutch-speaking subjects "diep na te denken."

II. RELIABILITY STUDIES

The remainder of this report describes a series of reliability studies. Beginning with the 120 stimulus pictures used in the pilot study, the authors successively screened them for stability and finally reduced them to an experimental set of 14. Subjects (judges) in these studies were college students who were tested during sessions of their introductory psychology laboratory classes. The stimulus pictures were presented to them in the form of 2 × 2 slides, which were shown one at a time, in a semidarkened classroom, by means of a remotely controlled projector. The screen image was 17 inches high by 25 inches wide; the first row of students was seated 6-1/2 feet from the screen.

A. Test-Retest I

Five of the six original picture sets and the "replication" set from the pilot study provided six lists of 20 slides each. The order of the 20 pictures was randomized within each set to provide a list. Two groups of judges were used, with $N = 22$ per group. One group was administered lists A, E, and C, then a number of interim tasks as described below, and then lists A, E, and C again but with the order of presentation reversed within each list. The other group went through the same procedure but with lists B, F, and D.

1. *Procedure*

S received a mimeographed booklet of seven pages, containing (*a*) a face page for identification data and space for five examples and (*b*) six additional pages, one page for each list of 20 slides. Instructions were:

"This experiment is concerned with how people observe things. We are going to show you some slides of people's faces. There will be six groups of these slides—and we will show the slides one at a time. Each slide will be a person's face—and you may notice some faces that were photographed more than once." Instructions were then read aloud. "For each slide, your task will be to judge the expression on the person's face. Please look at your top sheet. You will notice that there are two columns, marked with the words Thinking and Not Thinking. Now, when you look at a slide, you are to judge whether the person is or is not engaged in serious thinking or solving a problem.

"We will now show you the example slides—one at a time. You will mark your answers in the following way: If the person shown on the slide appears to *you* to be in the act of serious thinking, place an X or check in the column marked Thinking. If the person does not appear to *you* to be in the act of serious thinking, place an X or check in the column marked Not Thinking.

"Check the column that you think best describes the expression—and remember, even if you are not sure, ALWAYS CHECK ONE OF THE COLUMNS. Don't leave any out—even if you have to guess."

The examples, consisting of a number of neutral expressions, were then presented, and questions by *S*s were answered. They were allowed eight seconds per slide.

A 30-second pause was given between each list of slides. Following the presentation of the third list, the booklets were collected and *S*s were given a 10-minute break. They were then administered a number of interim tasks consisting of a group intelligence test and the Kuder Occupational Preference Record, Form D (13). No time limits were imposed, although all *S*s finished within 35 minutes. After the test material was collected, individual booklets were again distributed for recording judgments on the last three lists of slides. These slides were presented without reference to the fact that *S*s had viewed them previously.

2. *Results*

a. Nomenclature. Stimulus materials used in a social perception task may be categorized by the experimenter (*E*) or by the judges or perceivers (*S*s). The pictures utilized in the present study were initially categorized by *E*, on the basis of the situation in which they were produced, as either thinking (T) or nonthinking (N). On the basis of responses by *S*s in the test re-test procedure just described, each picture could now be additionally categorized as either T or N according to majority judgment by the *S*s of the reliability groups. For convenience, the letter R (for Reliability) was inserted as part of the label for each category, so that the four labels that could now be applied to a picture were TRT, referring to a picture categorized by *E* as thinking and also reliably judged by *S*s as thinking; TRN, a picture categorized by *E* as thinking but reliably judged by *S*s as nonthinking; NRN; and NRT.

b. Screening. By expansion of the binomial and reference to appropriate tables (25), the 120 pictures were scanned, and the 99 pictures which were scored as reliable at the .15 level or better were retained. For purposes of further screening, additional pictures were eliminated as follows: 1) whenever a picture had been used in more than one list, one of the pair of repeats was deleted; three pictures were deleted by this criterion; 2) all pictures originally in the "High Choice-Low Choice" set were deleted on the grounds that the choice among pictures of this set constituted a special task; 3) all pictures from list *E*, the "thinking/imitation of thinking" set, were deleted if they were pictures of imitations of thinking; those which had been experimentally arranged as thinking pictures were retained. The total retained was now 71 pictures. One "imitation" picture from list *E* was also included by error, making a total of 72 that were utilized in subsequent experiments.

By randomizing the order of these 72 pictures and dividing them into four lists of 18 pictures each, there were obtained lists AA, BB, CC, DD. These are presented in Table 1, together with the results for all the reliability studies reported here. Of the 72 pictures, 38 were of males, 20 of which had originally been classified by *E* as thinker and 18 as nonthinker; and 34 were of females, of which 17 had been classified by *E* as thinker and 17 as nonthinker.

B. JUDGMENTAL AGREEMENT

In addition to a measure of reliability based on repeated presentation of the stimulus material to the same *S*s, consistency of judgment may also be assessed in terms of agreement among *S*s—that is, by counting the number of *S*s who agree in judging a particular photograph. Two such "agreement" analyses were made of the data.

1. *Agreement I*

Based on the data from the two groups of *S*s in Test-Retest I and the results from the first ("Test") presentation of pictures, the probability was calculated for each of the pictures separately that the obtained agreement among judges was due to chance. The results for 72 selected pictures appear in Table 1, in the columns under Agreement I.

2. *Agreement II*

Data were also available for analysis of agreement among 102 additional *S*s who had judged the 120 pictures in a separate study. Since each of six groups of *S*s in this study judged a different list of 20 slides first, the

TABLE 1
RELIABILITY DATA ON THE 72 PICTURES

			Test-Retest				Agreement					
			I		II		I		II		III	
Pic. No.	Sex	Category	Result	Signif.	Result	Signif.	Result	Signif.	Result	Signif.	Result	Signif.
					List AA							
G-1	F	T	N	.02	T	.002		NS		NS	T	.01
H-1	M	N	N	.01	N	.002	N	.004		NS	N	.02
I-1	M	T	N	.01	T	.02	N	.004		NS		NS
J-1	F	N	T	.06	T	.002		NS	T	.17	T	.002
K-1	M	N	N	.15	T	.002	N	.13	T	.002	T	.01
L-1	M	T	N	.15		NS		NS		NS		NS
M-1*	F	T	N	.01	N	.002	N	.02	N	.14	N	.002
N-1	M	T	T	.01		NS	T	.004		NS		NS
O-1	M	N	T	.06	T	.002		NS	T	.002		NS
P-1*	M	N	T	.01	T	.002	T	.004	T	.09	T	.002
Q-1	F	N	T	.01	T	.01	T	.05	T	.002		NS
R-1*	F	T	T	.01	T	.002	T	.004	T	.002	T	.002
S-1*	F	T	T	.15	T	.002	T	.13	T	.004	T	.04
T-1	F	N	T	.01	N	.002	T	.004		NS	N	.002
U-1	M	T	N	.01[a] .02	N	.002	N	.004[a] NS	N N	.002[a] .02	N	.002
V-1	F	T	T	.01	N	.01	T	.05	N	.05		NS
W-1	M	N	N	.15		NS		NS	T	.05		NS
X-1	M	T	N	.01		NS	N	.13	T	.06		NS
					List BB							
J-2	F	N	T	.01		NS	T	.02		NS		NS
X-5	M	T	T	.01	T	.01		NS	T	.002		NS
H-2	M	N	T	.01	T	.002	T	.13		NS	T	.02
Y-1	F	N	T	.01	T	.01		NS	T	.09		NS
I-2*	M	T	T	.01	T	.002	T	.004	T	.17	T	.01
Z-1	F	N	T	.01	N	.002	T	.004	N	.05	N	.002
L-2	M	T	T	.01		NS	T	.004		NS		NS
N-2	M	T	N	.06	T	.002		NS	T	.09	T	.002
O-2	M	N	N	.02	N	.002		NS		NS	N	.01
P-2	M	N	T	.15	N	.002		NS	N	.004	N	.002
Q-2	F	N	T	.01		NS	T	.13		NS		NS
R-2	F	T	T	.15[a] .02		NS		NS		NS		NS
S-2*	F	T	T	.01	T	.002	T	.004	T	.002	T	.01
T-2	F	N	N	.02		NS		NS	N	.03		NS
U-2	M	T	T	.01	N	.04	T	.004	N	.02		NS
V-2	F	T	N	.01	T	.01	N	.004		NS	T	.04
W-2	M	N	T	.01		NS	T	.004		NS		NS
X-2	M	T	T	.01	N	.02	T	.004	N	.17		NS

* These pictures were selected for the Lexington Series.

[a] This picture appeared twice in the original list of 120 pictures; results for both presentations are reported.

TABLE 1 (*continued*)

			Test-Retest				Agreement					
			I		II		I		II		III	
Pic. No.	Sex	Category	Result	Signif.	Result	Signif.	Result	Signif.	Result	Signif.	Result	Signif.
					List CC							
N-5	M	T	T	.01	T	.002	T	.05		NS	T	.002
O-3*	M	N	T	.01	T	.002	T	.02	T	.01	T	.002
Z-2	F	N	T	.06	N	.002		NS		NS	N	.002
X-3	M	T	N	.02	T	.002		NS	T	.01	T	.002
V-3	F	T	T	.15	T	.002		NS		NS		NS
W-3	M	N	T	.01	T	.002	N T	.13[a] .004	T T	.004[a] .002	T	.002
T-3	F	N	T	.01	N	.01	T	.004		NS	N	.01
U-3	M	T	N	.01		NS	N	.004	N	.06		NS
R-3	F	T	T	.02	T	.002	T	.02[a] NS	T	.004[a] NS	T	.01
S-3	F	T	N	.01	T	.002	N	.05		NS	T	.01
P-3	M	N	T	.06		NS		NS	N	.17		NS
Q-3	F	N	T	.01		NS	T	.004	N	.06		NS
Y-2	F	N	T	.15	N	.002		NS	N	.004	N	.01
G-2	F	T	T	.01		NS	T	.02		NS		NS
J-3	F	N	N	.02	T	.002	N	.004	T	.09	T	.002
K-2	M	N	T	.01	N	.002	T	.05		NS		NS
N-3*	M	T	T	.15	T	.002	T	.13	T	.02	T	.04
I-3	M	T	T	.01	N	.04	T	.05		NS		NS
					List DD							
Y-3	F	N	T	.01	N	.002	T	.05	N	.002	N	.002
K-3	M	N	T	.01		NS	T	.13		NS		NS
N-4	M	T	T	.15	T	.002		NS	T	.004	T	.002
Z-3	F	N	T	.01	N	.002	T	.13	N	.06	N	.01
T-4	F	N	N	.01	T	.002	N	.004	T	.06	T	.002
Q-4	F	N	T	.01		NS	T T	.13[a] .004	N	.06[a] NS		NS
U-4	M	T	T	.01	T	.002	T	.004		NS	T	.01
M-2	F	T	T	.01	T	.04		NS	T	.02		NS
P-4	M	N	T	.02		NS	T	.13	T	.05		NS
X-4	M	T	T	.01	T	.002	T	.004		NS	T	.002
R-4*	F	T	T	.01	T	.002	T	.13	T	.01	T	.002
O-4	M	N	T	.01		NS	T	.004	N	.17		NS
V-4*	F	T	T	.01	T	.002	T	.004	T	.01	T	.002
S-4*	F	T	T	.01	T	.002	T	.02	T	.004	T	.04
H-3*	M	N	T	.01	T	.002	T	.004	T	.002	T	.002
W-4	M	N	T	.01	N	.01	T	.05	T	.02		NS
L-3*	M	T	T	.01	T	.002	T	.004	T	.002	T	.01
G-3*	F	T	N	.01	N	.002	N	.004	N	.14	N	.01

probabilities were calculated separately for each list. The results, tallied for each of 72 pictures, are given in Table 1 in the columns under Agreement II.

C. TEST-RETEST II

Sixty new subjects, 24 males and 36 females, were administered the 72 pictures comprising lists AA, BB, CC, and DD, utilizing the procedure and appropriately modified instructions of the Test-Retest I experiment. The lists were presented in the order AA, BB, CC, DD, with a 30-second pause between lists. Following the presentation of list DD, *S*s were given an interim task consisting of a 68-word mimeographed passage in Italian, which they were asked to copy by using their nonpreferred hand. *S*s were given 10 minutes for this task. All finished without difficulty in the time allowed.

Following the interim task, the four lists of pictures were presented again, in the order AA, BB, CC, DD, but with the order of presentation reversed within each list. A 30-second pause was given between lists.

The results for the 72 pictures are presented in Table 1, in columns "Result" and "Signif." under Test-Retest II.

D. JUDGMENTAL AGREEMENT III

A portion of the data from the above Test-Retest II task may also be analyzed in terms of agreement among judges. Responses to the first presentation only, preceding the interim task, were summed for the 60 *S*s. The results for this analysis of each of the 72 pictures are entered in Table 1, in the columns labeled "Result" and "Signif." under Agreement III.

Inspection of Table 1 indicates no difference between male and female photographs in regard to the reliability with which they may be judged. If there is any other basis for the fact that some expressions are judged with greater consistency than others, it will have to be sought in the influence of such variables as the person depicted, the specific expression assumed by the stimulus person, or an interaction effect. In any case, it is clear from inspection of the large number of reversals and changes as one reads across each line of Table 1, that stimuli of this sort, even when utilized in a straightforward dichotomous-choice task, may be quite unstable. Great care must be exercised in selecting such materials for experimental purposes, and results based on responses from only one set of judges are to be interpreted with some caution. There are, however, some pictures which seem to be stable enough to be retained as stimuli in a social perception task.

E. THE "LEXINGTON" SERIES

The results on five reliability measures (Table 1) were inspected and 14 photographs were selected for use in further studies; to these were added 11 other "control" photographs, one for each of the persons depicted in the set of 14, which were judged by *E* to be adequate representations of the target persons in a relaxed (nonthinking) pose. The 25 selected photographs, called the "Lexington Series," were randomly assigned three-digit numbers for identification, and appear to provide a stable set of stimulus materials for use in experiments on social perception.[6] The photographs, identified by an asterisk in Column 1 of Table 1, are presented in Figures 1, 2, 3, and 4.

---

[6] For information on the purchase of prints or slides of the Lexington Series, interested persons are invited to communicate with the senior author.

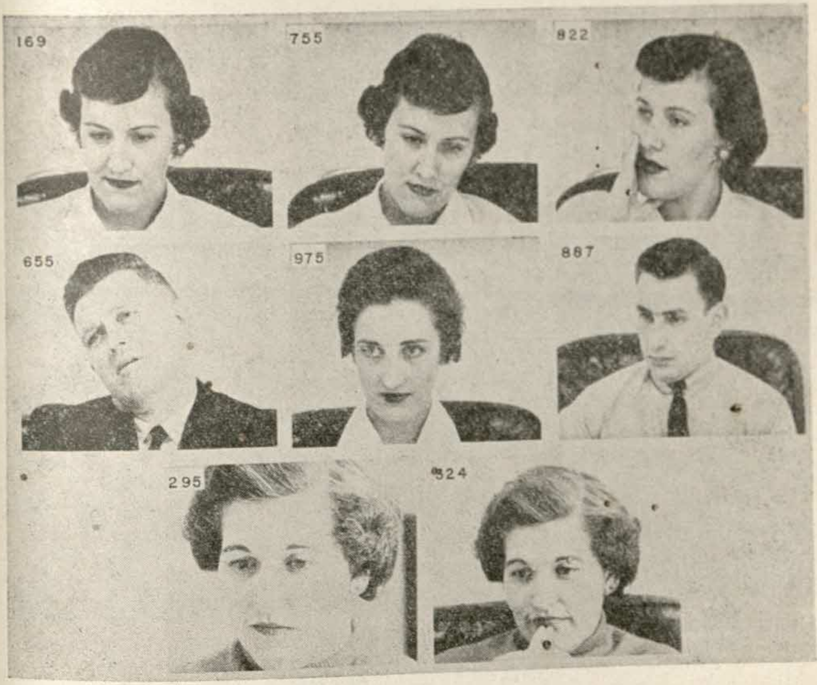

FIGURE 1
THE LEXINGTON SERIES—PICTURES WHICH WERE SELECTED BY THE RELIABILITY MEASURES DESCRIBED IN THE TEXT
(For data on the individual pictures, see Table 1)

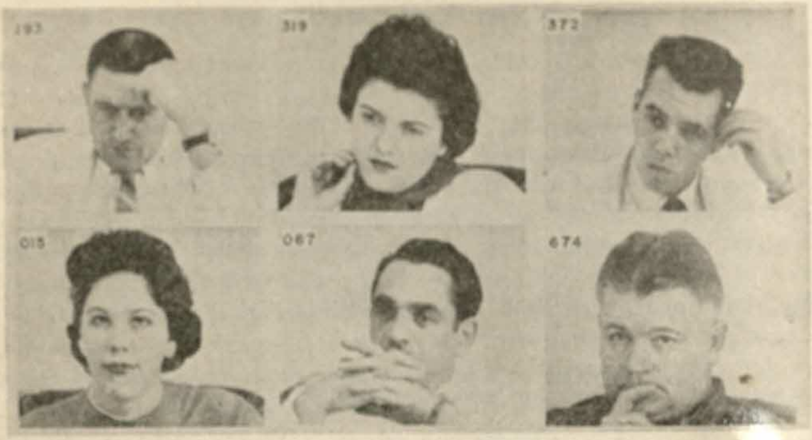

FIGURE 2
THE LEXINGTON SERIES—PICTURES WHICH WERE SELECTED BY THE RELIABILITY MEASURES DESCRIBED IN THE TEXT
(For data on the individual pictures, see Table 1)

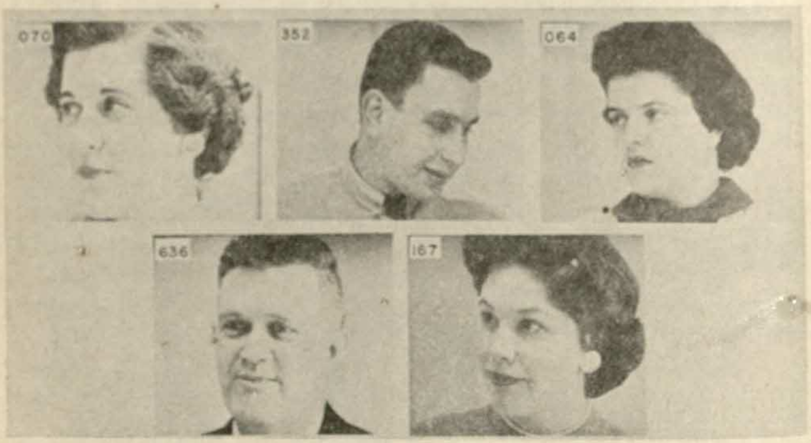

FIGURE 3
THE LEXINGTON SERIES—PICTURES OF THE TARGET PERSONS SMILING OR TALKING IN RELAXED FASHION

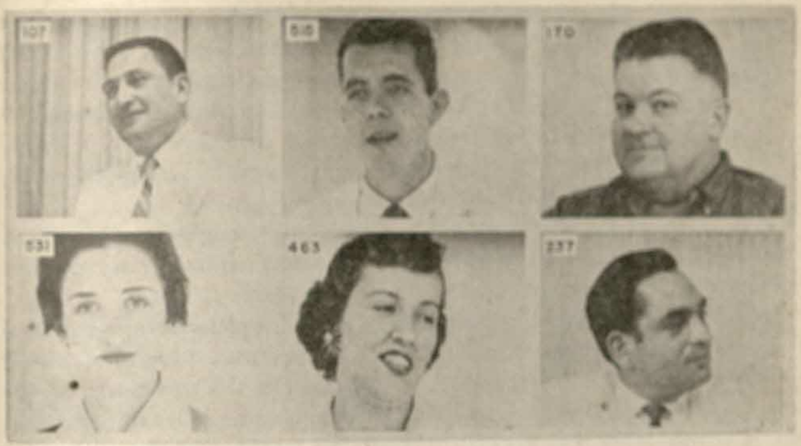

FIGURE 4
THE LEXINGTON SERIES—PICTURES OF THE TARGET PERSONS SMILING OR TALKING IN RELAXED FASHION

III. DISCUSSION

The pictures of the Lexington Series fall into three categories: (*a*) thinking pictures reliably judged as thinking, three male and six female—pictures numbered 655, 067, 887, 169, 755, 822, 324, 295, and 319; (*b*) thinking pictures reliably judged as nonthinking, two female—pictures numbered 975 and 015; (*c*) nonthinking pictures reliably judged as thinking, three male—pictures numbered 674, 372, and 193. Two females, both "thinkers," furnish more than one picture each. They are female S, with pictures numbered 169, 755, and 822; and female R, with pictures numbered 324 and 295. It is of some interest that, consistent with a (nonsignificant) steady trend throughout this research, pictures of females tend to be judged as nonthinkers and pictures of males as thinkers; and this trend finds support even in the case of the highly selected pictures of the Lexington Series.

The categories of pictures which "held up" on the basis of our multiple criteria of reliability suggest some tentative conclusions about the stimulus materials customarily utilized in social perception studies. Most of the pictures that turned out to be useable for the Lexington Series were of thinkers, and this fact alone indicates that facial expressions which are delimited and specified in advance may provide more stable stimuli than expressions selected on a more random basis. In this connection it should be noted that we were unable to find, for this series, any male TRN or female NRT pictures nor any male pictures in the NRN category which satisfied our criteria; here the sex of the picture interacts with the task and the categories given to the judges. It should also be noted that the experimental classification of a picture —that is, the initial selection by *E*—though useful initially as an aid, is hardly an infallible guide to its final status as a stimulus item. Five pictures in the Lexington Series, which were judged with great consistency through a series of tests, were classified differently by the *S*s than by *E*.

The results presented in the preceding sections, although promising, reveal a number of limitations inherent in our procedure and therefore must be interpreted with some caution.

It will be recalled that the "nonthinking" stimulus pictures were generated by placing behavers in situations deliberately devised so as to resemble closely situations of serious, directed thinking. As a result of this procedure, the two classes of pictures, thinking and nonthinking, were similar in expressive characteristics. They may have posed so subtle a task as to defeat some of the purposes of the study, especially in view of the fact that discrimination that is so refined is rarely found in nonlaboratory instances of social perception. Some of the results obtained—for example, discovering some NRT pictures which were consistently judged as thinkers although not originally "posed" as such—may be attributed in part to the close similarity between pictures originally selected as thinking and nonthinking. Thus, the gain in precision of results, which is accomplished by setting a specific task of discrimination to judges, is at least partially offset by the effect of unknown variables that may thus be introduced into the judging task. This limitation is equally pertinent in regard to the limited range of expression upon which the present study is based.

Another limitation in regard to the interpretation of our results concerns the test-retest procedures. We chose this as one method for assessing the reliability of our stimulus materials, since it is commonly used in experimental studies. In the initial reliability study (Test-Retest I), approximately one hour intervened between test and retest, during which an intelligence test and the Kuder Preference Record were administered. It was thought that this would be sufficient as an interpolated task to minimize the effects of memory from test to retest on the social perception tasks. In the second reliability study (Test-Retest II, Section II. C), only 10 minutes of interpolated activity were used between test and retest. Our results indicate that memory of the pictures, independent of any other factors, might have been an important influence in determining responses on the retest task. For this reason, some limitation is placed on assessing the value of our sequence of test-retest procedures for selecting reliable pictures. The authors would suggest, for this reason as well as on rational grounds, that assessment by means of consensus among judges provides a sound basis for such selection.

Given these limitations, however, it seems reasonable to conclude from our results, (*a*) that rigorous and objective studies of social perception may be undertaken in the laboratory; (*b*) that social perception judgments, at least in regard to the limited class of stimulus materials used here, do accord to a significant degree with criteria such as judgment by *E* and the nature of the situation; (*c*) that this gives us confidence in what is usually termed the "accuracy" of social perception by adults, particularly in view of the difficult task of discrimination posed to our judges; and (*d*) that appropriate procedures can be developed so as to produce reliable stimulus materials and tools upon which further experimental results can be based.

In regard to the last point, it is our expectation that the Lexington Series, when it is further refined and "purified," should provide more clear-cut results than have usually been obtained. Hopefully, we have taken at least the first step in what we have called the multi-stage, bootstrap operation that is required in experimental studies of social perception.

IV. SUMMARY

From motion picture records of 10 behavers in the act of serious, directed thinking, and of 10 other behavers in similar, nonthinking situations, 120 selected photographs were obtained and presented in sets of 20 for judgment by 118 *S*s in a dichotomous-choice task. Responses were evaluated in terms of agreement with choices by *E* and by peers. Marked differences were found among the stimulus pictures by either test-retest or consensus measures of reliability. Five reliability tests were utilized for successive screening, producing a set of 14 pictures felt to be stable enough to warrant further experimental use.

REFERENCES

1. BELL, C. Anatomy and Philosophy of Expression as Connected with the Fine Arts (7th ed., rev.). London: G. Bell & Sons, 1806.
2. BORING, E. G., & TITCHENER, E. B. A model for the demonstration of facial expression. *Amer. J. Psychol.*, 1923, **34**, 471-485.
3. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. Ramsey (Eds.), *Perception, an Approach to Personality*. New York: Ronald, 1951. Pp. 121-147.
4. COLEMAN, J. C. Facial expressions of emotion. *Psychol. Monog.*, 1949, **63** (Whole No. 296), pp. 36.
5. DARWIN, C. The Expression of the Emotions in Man and Animals. New York: Appleton, 1894.

6. DUNLAP, K. A project for investigating the facial signs of personality. *Amer. J. Psychol.*, 1927, **39**, 158-161.
7. EKMAN, P. Body position, facial expression and verbal behavior during interviews. *J. Abn. & Soc. Psychol.*, 1964, **68**, 295-301.
8. ENGEN, T., LEVY, N., & SCHLOSBERG, H. A new series of facial expressions. *Amer. Psychol.*, 1957, **12**, 264-266.
9. FELEKY, A. M. Feelings and Emotions. New York: Pioneer, 1922.
10. FROIS-WITTMAN, J. The judgment of facial expression. *J. Exper. Psychol.*, 1930, **13**, 113-151.
11. GUILFORD, J. P., & WILKE, M. A new model for the demonstration of facial expressions. *Amer. J. Psychol.*, 1930, **42**, 436-439.
12. HANAWALT, N. G. The role of the upper and lower parts of the face as a basis for judging facial expressions: II. In posed expressions and candid camera pictures. *J. Gen. Psychol.*, 1944, **31**, 23-36.
13. KUDER, G. F. Kuder Preference Record Occupational-Form D: Manual. Chicago: Sci. Res. Assoc., 1959.
14. LANDIS, C. Studies in emotional reactions. I. A preliminary study of facial expression. *J. Exper. Psychol.*, 1924, **7**, 325-341.
15. ———. The interpretation of facial expression of emotions. *J. Gen. Psychol.*, 1929, **2**, 59-72.
16. LANGFELD, H. S. The judgment of emotions from facial expressions. *J. Abn. & Soc. Psychol.*, 1918, **13**, 172-184.
17. LYONS, J. An experimental study of the expression of thinking. *Rev. Psicol. (Madrid)*, 1964, **19**, 17-28.
18. MUNN, N. L. The effect of the knowledge of the situation upon judgment of emotion from facial expressions. *J. Abn. & Soc. Psychol.*, 1940, **35**, 324-338.
19. PIDERIT, T. Wissenschaftliches System der Mimik und Physiognomik. Detmold: Meyer, 1867.
20. RUCKMICK, C. A. A preliminary study of the emotions. *Psychol. Monog.*, 1921, **30** (Whole No. 3), 30-35.
21. RUDOLPH, H. Der Ausdruck der Gemütsbewegungen des Menschen (textbook and atlas). Dresden: Kuhtmann, 1903.
22. SHERMAN, M. The differentiation of emotional responses in infants. I. Judgments of emotional responses from motion picture views and from actual observation. *J. Comp. Psychol.*, 1927, **7**, 265-284.
23. STRAUS, E. W. Rheoscopic studies of expression. *Amer. J. Psychiat.*, 1951, **108**, 439-443.
24. TOMKINS, S. S., & MCCARTER, R. What and where are the primary affects? Some evidence for a theory. *Percept. & Motor Skills*, 1964, **18**, 119-158.
25. U. S. Government. Tables of the Binomial Probability Distribution. Applied Mathematics Series 6, National Bureau of Standards, Washington, D. C., October, 1952.

Veterans Administration Hospital
Lexington, Kentucky 40507

Published as a separate and in *The Journal of Psychology*, 1966, 62, 83-87.

AN EVALUATION OF HUMAN PERFORMANCE DURING EXPOSURE TO ELEVATED HEAT AND HUMIDITY*[1]

Department of Anesthesiology, University of Washington School of Medicine, Seattle, Washington; and the United States Army Hospital, Fort Lawton, Washington

RICHARD J. WARD, EUGENE J. POWELL, MARK H. SCHORZMAN,
AND RONALD J. BENVENISTE

A. INTRODUCTION

Man's reaction to heat and stress has been studied with respect to some of the changes in body physiology and performance. Lind, reporting on body temperature changes during work performed in an elevated environmental temperature, noted that, for a given rate of work, the rectal temperature is practically independent of the influence of environmental temperature (4). After the first hour of work in an elevated temperature, there were no changes in temperature, pulse rate, or body weight (5).

There is, however, an environmental temperature at which decompensation occurs. Increased humidity seems to hasten this decompensation. This reaction is vividly demonstrated in the summer and fall of each year. Deaths continue to be reported following athletic practice or games played in hot, humid weather. The Armed Forces have issued directives prohibiting calisthenics and outdoor training when the temperature and humidity reach certain maxima.[2]

Changes in blood pressure, pulse rate, rectal temperature, and other similar parameters under conditions of adverse temperature and humidity do not reflect the performance of man as a whole. It is a common experience to notice that working in a hot or humid climate seems to make one work less efficiently. Yet performance under these conditions does not seem to have been adequately evaluated.

A method of performance determination that has been employed to evaluate other stresses is that of flicker-fusion frequency, which may be defined as "that rate of successive light flashes from a stationary light source at which the sensation of flicker disappears and the light becomes steady" (9). This method

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 10, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The information contained herein is the personal opinion of the authors and does not necessarily reflect the policy of the Department of the Army.

[2] Tech. Bull. Med. 175, Dept. of the Army.

of performance testing has many advantages: namely, simplicity, short duration of test application, lack of practice trends, and sensitivity to a variety of physiologic stress situations, such as hypoxia, fatigue, and anxiety. It is a measure of the brain's ability to identify, integrate, and respond to a stimulus. As such it is independent of the *IQ* (7). In this study flicker fusion was used to determine the change in human performance during exposure to heat, humidity, and mental stress.

B. Method

Flicker can be obtained either electronically or by interrupting a beam of light by a perforated rotating disc accompanied by film projection. Electronic methods were used in this study because of the ease of operation and absence of noise and vibration.

The flicker-fusion frequency was determined by increasing the rate of flashes in steps until the sensation of fusion was produced. This method of discontinuous exposure has been shown to be preferable to the gradual change over a prolonged exposure because of the dependency of flicker-fusion frequency on acceleration speeds in the latter technique (9).

Forty-four soldiers were used as subjects. They were told that they had been chosen to participate in an important study. The control determination of the flicker-fusion frequency was taken at 76° F and 40 per cent humidity. Twenty-nine of the subjects were then placed in a room at 82° F and 70 per cent humidity for three hours. For the first two hours they viewed general interest movies, following which they were subdivided into two groups. The members of Group I were told that they would have to take a written examination. This was a difficult civics test. It was mentioned that the test would be very important to them personally, and Viet-Nam was mentioned. Therefore, Group I was placed under conditions that were mentally stressful and accompanied by elevated temperature and humidity. The other group, Group II, was required to stay in the room for the designated time. Group II was exposed only to the elevated temperature and humidity. At the end of the three hours flicker-fusion frequency was again determined in these two groups.

The remaining 15 soldiers were designated as Group III. The members of this group were given identical instructions followed by the same difficult civics examination as that given to Group I. However, this was done under identical conditions of comfortable temperature and humidity as when the control flicker-fusion frequency was determined (75° F and 40 per cent relative humidity). After the test, the flicker-fusion frequency was once again determined.

In this study each subject served as his own control. After the final flicker-fusion frequency was determined, the purpose of the study was explained to the participants.

C. Results

Group I showed a mean decrease in flicker-fusion frequency of 2.03 flashes per second (fps) from that of the control ($p < .01$).

Group II showed a mean decrease of 1.29 fps, which was also significant ($p < .01$).

Group III showed a mean decrease of .57 fps ($p < .05$).

Group I differs statistically from Groups II and III ($p < .01$).

D. Discussion

The currently accepted hypothesis is that the flicker-fusion frequency is related to the excitability of the central nervous system (9). This is supported by studies with various drugs, physiologic stresses, and disease conditions. Robak, Krasno and Ivy showed that 96 mg of secobarbital lowered the flicker-fusion frequency by 2.0 fps in 15 subjects (8). Landis and Zubin demonstrated that 150 mg of phenobarbital lowered the flicker-fusion frequency by .6 fps in 24 subjects (2). These values are useful for comparison because of the commonly accepted depressant action of these drugs.

Group I, by showing a decrease of 2.03 fps in flicker-fusion frequency, demonstrated that experimental conditions for this group produced a depressant action equivalent to 96 mg of secobarbital, and considerably more than the action of 150 mg of phenobarbital. Group II, though demonstrating a slightly lesser degree of decrease in flicker-fusion frequency than Group I, nevertheless showed considerably more depressant action than that resulting from 150 mg of phenobarbital. Group III showed results equivalent to this dosage of phenobarbital.

The combination of mental stress and increased environmental heat and humidity has an adverse effect on performance. The adverse effects of elevated environmental heat and humidity are demonstrable, but of less clinical significance. In contrast to the findings of Simonson and Enzer, our subjects had a clinically insignificant decrease in function following a relatively short (one hour) period of severe mental stress (10).

Taylor has postulated a theory that anxiety leads to a facilitation of performance (11). However, Wagoner, using flicker-fusion frequency, showed this not to be the case (12). Other studies have been performed in an attempt to relate mental stress to flicker-fusion frequency. A study done on pilot instructors showed no change during the working day when no high altitude flying was involved (1). However, various forms of mental work, such as microscopy, laboratory, and clerical work, have been shown to result in a decrease in flicker-fusion frequency of 3 to 7 fps (10). Lee has shown that automobile driving results in a drop of 1.3 fps (3).

The phenomenon of fatigue should probably be considered, both in the aspects of mental and physiologic stress. However, fatigue is complex to evaluate by flicker fusion, perhaps because of the possible interference of a central feed-back mechanism (6). Studies with sleep deprivation by different workers have shown different and opposite results (9).

The results of this study would indicate that working under conditions of stress and increased temperature and humidity is detrimental to performance. The temperature and humidity ranges studied were not excessive but are commonly observed in many parts of our country. The depression of responses under these conditions approaches or may even exceed those of depressant drugs. Though stress and anxiety may be difficult to eliminate, temperature and humidity can be controlled.

E. Summary

Forty-four soldiers participated in an evaluation of performance during exposure to increased heat and humidity. Determination of performance capability was made by measuring the flicker-fusion frequency (FFF) of each subject. FFF was chosen because it is an accurately measurable indicator of cerebral function. It is independent of the *IQ*.

Subjects were initially tested under comfortable environment conditions. Group I (15 subjects) and Group II (14 subjects) were placed for three hours in a room with a temperature of 82° F and 70 per cent relative humidity. At the end of two hours, Group I subjects were given an unannounced difficult civics test and told that the results of the test would be important to them. At the end of the third hour, the final FFF was measured in both groups. Group III was given the same unannounced civics test, but in a room with normal temperature and humidity (74° F and 40 per cent humidity).

Group I had a mean FFF decrease of 2.03 fps. Group II had a mean FFF decrease of 1.29 fps. Group III had a mean FFF decrease of .57 fps.

It is concluded that one's mental capabilities are blunted by mental stress and hot, humid environmental conditions. This blunting exceeds that found

after medication with commonly prescribed soporifics. The adverse environmental conditions themselves caused moderate depression of function. Mental stress alone caused clinically insignificant depression of function.

REFERENCES

1. GREYBIEL, A., LIBENTHAL, J. L., JR., & HORWITZ, D. Flicker fusion tests as a measure of fatigue in aviators. *J. Aviat. Med.*, 1943, **14**, 356-359.
2. LANDIS, C., & ZABIN, J. Effects of thomgylamine hydrochloride and phenobarbital sodium on certain physiological functions. *J. of Psychol.*, 1951, **31**, 181-200.
3. LEE, R. H. Fatigue and hours of service of interstate truck drivers. Critical fusion frequency of flicker. *Pub. Health Bull.*, 1941, **265**, 195-208.
4. LIND, A. R. A physiologic criterion for setting thermal environmental limits for everyday work. *J. Appl. Physiol.*, 1963, **18**, 51-56.
5. ———. Physiological effects of continuous or intermittent work in the heat. *J. Appl. Physiol.*, 1963, **18**, 57-60.
6. LOWENSTEIN, O. Disintegration of central autonomic regulation during fatigue and its reintegration by psychosensory controlling mechanisms. *J. Nerv. & Ment. Dis.*, 1952, **115**, 1-21, 121-145.
7. NELSON, T. M., & BARTLEY, S. H. Theoretical interpretations of various qualitative and quantitative aspects of flicker and fusion phenomena. *J. of Psychol.*, 1965, **59**, 185-194.
8. ROBACK, G. S., KRASNO, L. R., & IVY, A. C. Drug effects on flicker fusion threshold. *J. Appl. Physiol.*, 1952, **4**, 566-574.
9. SIMONSON, E., & BROZEK, J. Flicker fusion frequency: Background and applications. *Physiol. Rev.*, 1952, **32**, 349-378.
10. SIMONSON, E., & ENZER, N. Measurements of fusion frequency of flicker as a test of fatigue of the central nervous system; observations on laboratory technicians and office workers. *J. Indus. Hyg. & Toxicol.*, 1941, **23**, 83-89.
11. TAYLOR, J. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303-320.
12. WAGONER, R. A. Differences in response latency and response variability between high and low anxiety subjects in a flicker fusion task. *J. Abn. & Soc. Psychol.*, 1960, **61**, 355-359.

Department of Anesthesiology
University of Washington School of Medicine
Seattle, Washington

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 89-97.

EFFECTS OF CONGENITAL FELINE CEREBELLAR HYPOPLASIA*[1] ON DEVELOPMENTAL BEHAVIOR AND THE VESTIBULAR SYSTEM

Vestibular Section, Aeromedical Indoctrination Branch, USAF School of Aerospace Medicine, Brooks AFB, Texas

P. J. Dowd

A. Introduction

Throughout the literature concerning congenital feline cerebellar hypoplasia, a few cases have been reported (2, 3, 4, 7, 16). It appears that this congenital defect is particularly prevalent in cats. This report is concerned with the behavioral development of six cats with cerebellar hypoplasia and of their vestibular responses to rotary stimulation. The role of the cerebellum as the great coordinating center was studied in relation to the physiologic consequences of pathologic interference with the cerebellar afferent and efferent systems. A perspective of their behavioral development and physiologic functions, a description of their structural defects, and the results of their vestibular responses to sinusoidal accelerations will be discussed with this objective in mind.

B. Developmental Behavior

Four of the six cats were from one litter. Of the remaining two cats, one was from two surviving kittens of a different litter, and the other one was from an unknown source. All six cats were received between the ages of two and six weeks at this laboratory. In each case observations were possible for considerable periods of time, which covered one cat for six months, one for eight months, two for two years, one for three years, and one presently nearing its fifth year. Throughout their development and the previously stated periods of observation, their temperament was docile and they were easy to handle, feed, and maintain. No problems existed in their appetites, feeding, or swallowing. They were quite active, playful, attentive to their cleanliness, and they purred and mewed like normal cats.

---

* Accepted by Heinrich Klüver of the Editorial Board, received in the Editorial Office, Provincetown, Massachusetts, on January 10, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This work was carried out at the USAF School of Aerospace Medicine, Brooks Air Force Base, Texas. Appreciation is extended to Dr. G. L. Anstadt and Dr. J. C. Brown for their assistance and advice in animal care. The study reported herein was conducted according to the "Principles of Laboratory Animal Care" established by the National Society for Medical Research.

Ataxia was the main and obvious result of cerebellar hypoplasia. Except for the latter, the general developmental condition and behavior of these six cats were no different from normal cats in our laboratory. Their respiratory, cardiovascular, digestive, and urogenital systems appeared to be normal. The musculoskeletal formations were normal with good muscle tonus and no lameness. During the first four months of their development it appeared that ataxia severely increased and inhibited any locomotion. After four months, they seemed to accommodate to their problem and increased their movements. Two of the six cats are presently being studied. The other four have expired. As a result of feline pneumonitis, two expired at the age of two years. One three-year-old died from an overdose of anesthesia during surgery for implanting chronic electrodes (9). A six-month-old cat was sacrificed after vestibular testing for histological studies.

C. Postural and Progressive Movements

When the animals attempted to stand, they arched their backs, stood on a broad base, using their hind legs in exaggerated supporting reactions, and displayed severe oscillating movements of the head, neck, and trunk. Figure 1 shows a good example of this standing position. Such movements were severe when the animals fixated on objects, as in attempting to pick up a piece of food. A constant tremor of the head, neck, and shoulders sometimes occurred while the animals were standing. The lower dorsal and lumbar spinal areas displayed an arched position. A "hammer" movement of the head, neck, and trunk occurred as the animals lowered their heads to lap milk or eat food from a saucer, which movement Brouwer (2) regarded as "hypermetria of neck musculature." The animals never stood up on their hind legs without external support.

In walking, the animals staggered, and incoordination was especially noticeable as they fell in either direction. Asynergic movement of the hind legs was manifested during walking by a widespread stance in which the legs were everted to control balance. There appeared to be a lack of coordination between hindleg and foreleg movements in progressive motion, in which the shoulder girdle remained rather stable while the pelvic girdle and appendages overrode the foreleg movements, causing side-to-side oscillations which ended by stumbling. Despite these asynergic movements of the hind legs, the animals made attempts to run which resulted in a "zig-zagging" course. In making

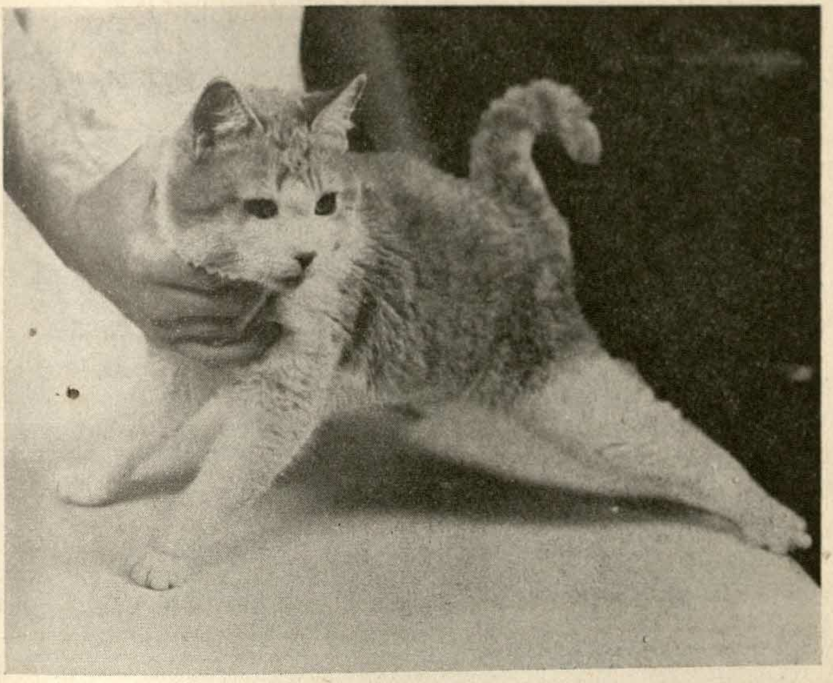

FIGURE 1
A SIDE VIEW OF A FOUR-YEAR-OLD CEREBELLAR HYPOPLASTIC CAT
STANDING ON A BROAD AND SECURE BASE
(His head is held to avoid head, neck, and shoulder movements)

turns while moving forward, they frequently fell or rolled over. Climbing or jumping actions were quite awkward and at times impossible.

D. REFLEXES

Upon examination, no paresis of limbs, nor asthenia, nor hypotonicity of muscles was found. Tendon reflexes were not altered and no myotonia was observed. Manipulation of forelimbs was quite precise when the animals were striking at moving objects while they lay on their side or in the prone position. Knee-jerk reflexes were active and no clonus was observed. Tail reactions while the animals were moving were normal. Labyrinthine-righting and neck-righting reflexes were present. Their body-righting reflexes appeared to be weaker than normal. Head and eye movements appeared normal. When the animals shook their heads, however, they would do a backflip, falling on either side. This behavior indicates a lack of cerebellar control on the activity of the neck musculature, initiated through the first few pairs of the dorsal cervical spinal roots. Limb reflexes and crossed extension reflexes of leg tendons appeared normal.

When blindfolded, the animals showed severe deficiencies in placing reactions, in labyrinthine-righting reflexes, and in body-righting reflexes. For example, when the blindfolded animals were dropped from a height of two or four feet onto a sponge rubber pad, they all usually fell on their sides. With a visual reference, however, the animals usually showed normal righting reflexes in that they would alight on all fours, though, at times they would fall on their sides. Also, the blindfolded animals could not place their forepaws on a tray when the dorsal area of their forepaws touched the edge of the tray, nor when their chins touched the tray. Their dependency on visual information was quite evident in that when they were blindfolded they would never stand or attempt to walk and they usually remained in the prone or recumbent position.

E. Sensory Effects

Sensory systems appeared normal in that the animals responded to or were sensitive to touch, and to hot and cold stimuli, and reacted to pinching and to stroking of the vibrissae. They sneezed when their nostril hairs were stroked. Their eyes responded normally to light and touch, and they blinked to air puffs. No spontaneous nystagmus was observed for all six cats. Sensations of taste and smell appeared normal. Normal reactions to pain were evident, contrary to Cobb's work (4). During the first few months of development the animals squealed when hitting their heads against the cage wall or the food saucer. They seemed to enjoy being petted and touched. Attempts to mate a male cerebellar hypoplastic cat with a female cerebellar hypoplastic cat failed. It was impossible to induce oestrus in the female.

F. Cerebellar Defects

In Figure 2 the size of the cerebellum of a cerebellar hypoplastic cat is compared with a normal-sized cerebellum of a cat. On macroscopic examination of the CNS structures of the cerebellar hypoplastic cats it was apparent that cerebellum sizes were reduced to about one-third of the size for a normal cat. The underdeveloped cerebellum on the right in Figure 2 illustrates the exposure of the inferior colliculi rostrally and also the posterior medullary velum caudally. This was true for all three cats examined. This symmetrical reduction, with a well-preserved general contour of the cerebellum

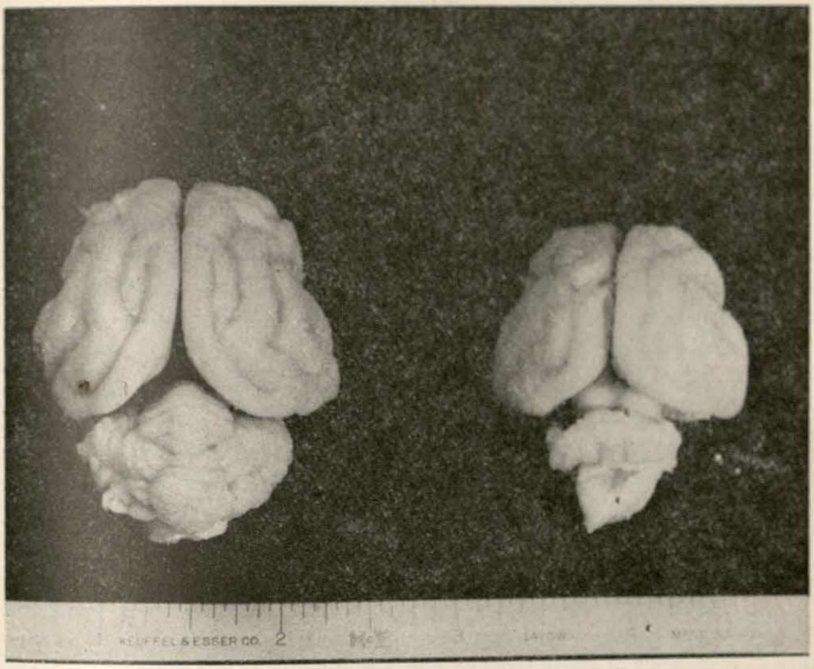

FIGURE 2
DORSAL SURFACE OF TWO CAT BRAINS
On the left is a normal sized cerebellum of a two-year-old cat. On the right is the marked defective cerebellum of a six-month-old cerebellar hypoplastic cat.

with similar differentiated convolutions and sulci as in normally developed cerebellar structures, was found for three cats examined. Forebrains appeared to be of normal size and the skull had no abnormal features and no asymmetry was noted in the skull structures. Two of the six cats are still being tested and another was unfortunately not saved for postmortem examination. Histological analyses are being presently made on the six-month-old and the three-year-old cats.

G. VESTIBULAR TESTING AND NYSTAGMUS

Two of the six cerebellar hypoplastic cats were tested on the School of Aerospace Medicine (SAM) Precisional Rotational Stimulator. One cat was a six-month-old male and the other was a four-year-old male. Two normal two-year-old males were used as controls. The stimuli presented on this yaw-axis rotating chair appear in Table 1. Electronystagmograms were recorded

TABLE 1
STIMULUS PRESENTATION

Test	Frequency (cps)	Peak acceleration (rad/sec^2)
1	.0400	.1
2	.0150	.1
3	.0225	.2
4	.0400	.4

of both cerebellar hypoplastic cats and compared with the ENG records of normal cats to the same stimuli. Examples of these two cerebellar hypoplastic cats' responses to stimulus 1 (.04 cps, .1 rad/sec^2) are compared with two normal cats' responses in Figure 3.

The vestibular nystagmic responses of the cerebellar hypoplastic cats elicited by rotary stimulation were qualitatively similar to normal cats' responses to the same stimuli, in that alternating slow and fast phases with the slow component of nystagmus coinciding with the rotary stimulation were recorded (8). Thus, despite the defective cerebellar structures of these cats, they showed no marked nor significant changes in their sensitivity, in the direction of the slow phase of their nystagmus, or in phase modifications of nystagmus from the classical examples of normal animals. As a result, the role of the cerebellum as one regulating or modifying vestibular input is questioned. It appears that the above-mentioned findings on vestibular responses to rotary stimulation support other results in the studies of De Kleyn (6), Dowd (8), Lorente de No (14), and Spiegel and Price (17), which indicated that the vestibular nuclei in the brain stem have an integrative as well as a relay function.

It was also noted that three of the six cerebellar hypoplastic cats, while being anesthetized with sodium pentobarbital, showed not only the usual and normal reactions of nystagmus, but also head-righting reactions. This further indicates the dependency of the neck reflexes on cerebellar control.

H. DISCUSSION

Further information concerning the histology and the quantitative analyses of two cerebellar hypoplastic cats' responses to rotary stimulation is in preparation. At present, the almost five-year-old cat is being studied for the role of habituation to vestibular stimulation by Fernandez at the University of Chicago. Halstead *et al.* (11, 12) found that lesions in the cerebellar areas, especially the nodulus, prevented the occurrence of habituation to vestibular

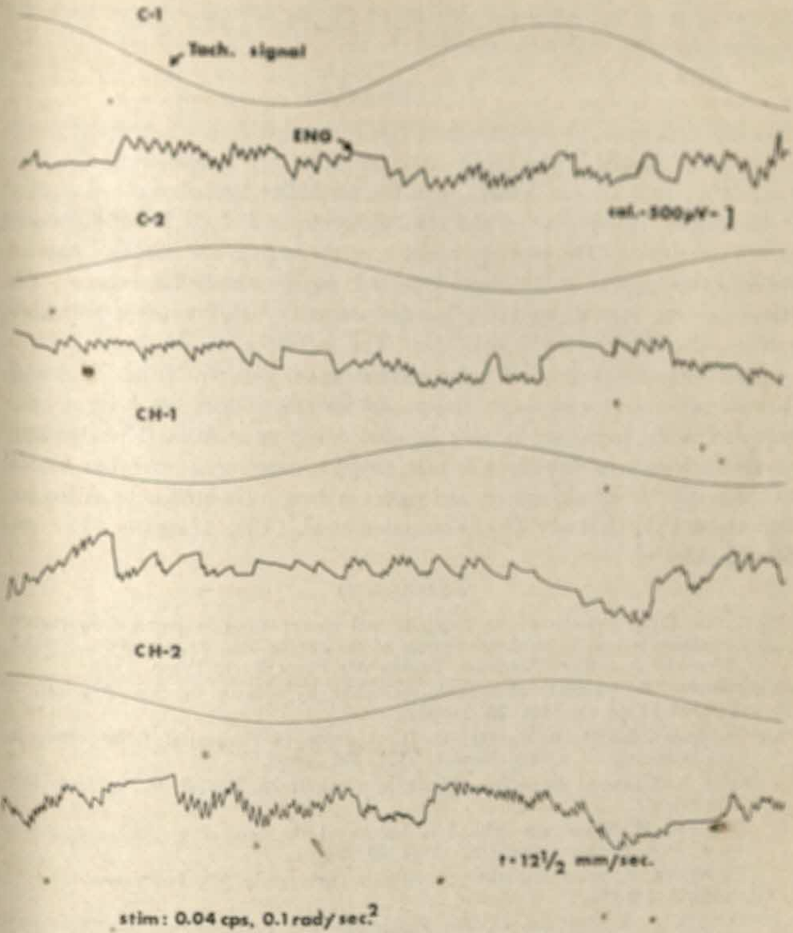

FIGURE 3
Examples of Nystagmic Responses of Both Normal (C-1, C-2) and Cerebellar Hypoplastic (CH-1, CH-2) Cats to a Sinusoidal Stimulus of a Frequency of .04 CPS and a Peak Acceleration of .1 Radians Per Second Squared

stimulation. Fernandez's study should reveal further concepts on cerebellar control of the vestibulo-ocular reflex arc system.

Very little has been published concerning the etiology of this disease, and the cause of these familial cerebellar disorders is not known. Recently, an interesting study has suggested that ataxia in cats accompanied by cerebellar hypoplasia may be of viral origin (13).

I. SUMMARY

Emphasis is made of the important role of the brain-stem vestibular nuclei in the interrelations of the vestibulo-ocular reflex arc system when the cerebellar structures are not intact, as in the cerebellar hypoplastic cats studied in this report. It appears that the vestibulo-ocular reflex arc is not dependent on the cerebellum. The gravity receptors of the utricle and saccule—namely, the otoliths—appear to be dependent and require cerebellar control. The latter concept was demostrated by the animals' inability to correct their position in falling when blindfolded. The influence of the cerebellum in equalizing the tonus of lower reflex mechanisms is apparent from this report. The role of vision for placing reactions and for labyrinthine and body-righting reflexes is quite important in such familial defective animals. These findings suggest necessary modifications of past concepts concerning cerebellar control of the vestibular sensory system and motor system, as in studies by Allen and Fernandez (1), Cranmer (5), Fernandez *et al.* (10), Margaria (15), and Strong (18).

REFERENCES

1. ALLEN, G., & FERNANDEZ, C. Experimental observations in postural nystagmus, extensive lesions in posterior vermis of the cerebellum. Report 59-74, U.S.A.F. School of Aerospace Medicine, Brooks Air Force Base, Texas, 1959.
2. BROUWER, B. Familial olivo-ponto-cerebellar hypoplasia in cats. *Psychiat. en Neurol.* (*Bladen*), 1934, **38**, 352-367.
3. CARPENTER, M. B., & HARTER, D. H. A study of congenital feline cerebellar malformations. *J. Comp. Neurol.*, 1956, **105**, 51-93.
4. COBB, S. A case of cerebellar aplasia in a cat. *Arch. Neurol. & Psychiat.*, 1928, **19**, 931-932.
5. CRANMER, R. Nystagmus related to lesions of the central vestibular apparatus and the cerebellum. *Ann. Otol.*, 1951, **60**, 186-196.
6. DE KLEYN, A. Some remarks on vestibular nystagmus. *Confin. Neurol.* (*Basel*), 1939, **2**, 257-292.
7. DOW, R. S., & MORUZZI, G. The Physiology and Pathology of the Cerebellum. Minneapolis, Minn.: Univ. of Minn. Press, 1958. Pp. 438.
8. DOWD, P. J. Responses of the habituated vestibulo-ocular reflex arc to drug stress. Report 64-72, U.S.A.F. School of Aerospace Medicine, Brooks Air Force Base, Texas, 1964.
9. DOWD, P. J., ANSTADT, G. L., & KOEGEL, E. A new chronically implanted electrode for recording nystagmus in animal. Report 62-125, U.S.A.F. School of Aerospace Medicine, Brooks Air Force Base, Texas, 1962.
10. FERNANDEZ, C., LINDSAY, J. R., & ALZATE, R. Postural nystagmus due to localized cerebellar lesions in the cat. *Acta Otolaryng.* (*Stockholm*), 1959, **50**, 287-291.
11. HALSTEAD, W. The Effects of Cerebellar Lesions upon the Habituation of Post-rotational Nystagmus. Evanston, Ill.: Northwestern Univ. Press, 1935.

12. HALSTEAD, W., YACORZYNSKI, G., & FEARING, F. Further evidence of cerebellar influence in the habituation of after-nystagmus in pigeons. *Amer. J. Physiol.,* 1937, **120**, 350-355.

13. KILHAM, L., & MARGOLIS, G. Cerebellar disease in cats induced by inoculation of rat virus. *Science,* 1965, **148**, 244-246.

14. LORENTE DE NO, R. Vestibulo-ocular reflex arc. *Arch. Neurol. & Psychiat.,* 1933, **30**, 245-291.

15. MARGARIA, R. Wide range investigations of acceleration in man and animals. *J. Aviat. Med.,* 1958, **29**, 855-871.

16. RUSSELL, J. S. R. Defective development of the cerebellum in a puppy. *Brain,* 1895, **18**, 523-530.

17. SPIEGEL, E. A., & PRICE, J. B. Origin of the quick component of labyrinthine nystagmus. *Arch. Otolaryng.,* 1939, **30**, 576-588.

18. STRONG, O. S. Unsolved problems suggested by cerebellar connections and cerebellar histology. *Arch. Neurol. & Psychiat.,* 1928, **19**, 1-30.

Vestibular Section, SMTI
USAF School of Aerospace Medicine
Brooks Air Force Base, Texas

Published as a separate and in *The Journal of Psychology*, 1966, 62, 99-110.

CHANGES IN CONFORMITY BEHAVIOR ON A PERCEPTUAL TASK FOLLOWING VERBAL REINFORCEMENT AND PUNISHMENT*[1]

Institute of Physical Medicine and Rehabilitation, New York University Medical Center

ROBERT GOODKIN[2]

A. INTRODUCTION

Asch (1, 2), Blake (5), and others found that a small group instructed to respond incorrectly to a perceptual task succeeded in deflecting about one-third of the responses of true subjects in the group. However, there were great individual differences in the number of conforming responses given, some subjects reporting completely independently from the majority, others completely agreeing with it. The present study was designed to investigate the effects of verbal reinforcement and punishment on an individual's responses to such a task.

The task of present concern was to have the subject report the number of clicks presented on a metronome after four other "subjects" had responded. Actually, the other "subjects" were tape-recorded graduate students reporting an incorrect number of clicks. Blake, Helson, and Mouton (5) found this tape-recording method to be successful in a very similar task. The percentage of responses deflected by the majority was in keeping with previous studies in which all of the group members were physically present.

The great interest that has developed in verbal conditioning is shown in three reviews of the literature, one by Krasner (16), one by Salzinger (18), and one by Greenspoon (11). The initial verbal conditioning research was done by Greenspoon (8, 9, 10). In this work he created an experimental situation to study verbal behavior that paralleled operant conditioning with infrahumans. The verbal stimulus "mm-hmm," a visual stimulus, and an auditory stimulus resulted in a significant increase in the response class which was the number of plural noun responses emitted by subjects in treatment groups when compared to a control group.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 11, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to thank Harold F. Cottingham, Jack Hokanson, and Willard Nelson for their helpful suggestions and criticisms.

[2] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Florida State University, April, 1965.

Buss and Buss (6) conducted a series of concept formation experiments in which "right" and "wrong" were manipulated together and separately. They found that "wrong" for the incorrect response and saying nothing for the correct response were more conducive to learning the correct response than saying "right" for the correct response and nothing for the incorrect one. Kirman (15) reported a decrement in responses followed by "not so good."

Several experimenters have investigated the effects of various schedules of reinforcement on the acquisition of a verbal response. Among these are Kanfer (13, 14), Grossberg (12), McNair (17), Solley and Santos (20), and Timmons (21). In general, schedules of reinforcement seem to affect the acquisition of a verbal response in the same manner which they affect other kinds of responses and responses in other organisms.

Several experimenters have investigated the effects of schedules of reinforcement on resistance to extinction. Kanfer (13) found that partial reinforcement schedules produced the most resistance to extinction. The continuous reinforcement group had the highest rate of responding at the start of the extinction period, but its rate dropped sharply. Fattu and Mech (7) and Grossberg (12) also reported that partial reinforcement resulted in increased resistance to extinction as compared with a continually reinforced group. In a later study, however, Kanfer (14) found that subjects reinforced on a fixed ratio, fixed interval, or variable interval schedule did not show signs of extinction although the fixed ratio groups gave the most responses when reinforcement was omitted.

Very little research has been done relating verbal conditioning to a group setting. An initial effort was made in this area by Bachrach, Candland, and Gibson (3), who created groups of three people, only one of whom was an actual subject. The other two group members were experimenters although the subject did not know it. The group's task was to decide the meaning of an ideogram. During an operant period, the experimenters talked about the meaning of the ideogram for a predetermined amount of time. The experimenters verbally reinforced the subject for talking during the conditioning phase. In the extinction phase, the experimenters omitted reinforcement of verbalizations by the subject by changing the topic of conversation proposed by the subject. In a reacquisition period, the same procedure was used as in the initial acquisition period. Although the results yielded considerable variability, a significant trend was demonstrated. There was a marked increase in the amount of time spent in talking by the subject during the acquisition and reacquisition phases. During extinction there was a sharp decline in the talking time of the subjects.

The present study investigates the effects of four schedules of verbal reinforcement given to subjects while responding to a task individually, on their responses to that same task when responding after four unseen "subjects" reporting incorrect responses to the task on the majority of the trials.

B. METHOD

1. *Subjects*

The subjects were 75 undergraduate students at Florida State University enrolled in the introductory psychology course. Subjects were assigned to one of five groups (Continuous Reinforcement Group, Partial Reinforcement Group, Continuous Punishment Group, Partial Punishment Group, Control Group), each group containing 15 subjects. Subjects were assigned to a group on the basis of the time order in which they participated. That is, the first 15 subjects to participate in the experiment constituted the first group, the next 15 subjects made up the second, and so on.

2. *Apparatus*

The experiment was conducted in a small room in the Counseling Laboratory at Florida State University. Subjects were seated at a table facing a one-way vision screen. The apparatus employed was a modification of that described by Blake and Brehm (4). Each subject heard the task and four background voices by means of a speaker resting on the table in front of him, and was told that the other group members were in adjoining rooms. Actually the task, consisting of series of clicks from a metronome (working at a rate of 200 beats per minute), and the responses of the four other individuals, two male and two female graduate students, were recorded on a stimulus tape, in order to provide the same task and responses by the background voices for each subject. A small microphone was placed next to the speaker on the table.

3. *Procedure*

Each subject was dealt with individually. After entering the experimental room, the subject was seated at the table and told that he was subject number five in the experiment. The experimenter then left the room and went into the adjoining observation room where he was able to view the subject through a one-way vision screen.

After five minutes, the tape recorder was started and the stimulus tape began as follows:

> Hello. I'm speaking from the control room. Before we get started, I would just like to make sure you can all hear me. So, when I call your number, if you can hear me, please say "OK." Number 1. (OK.) Number 2. (OK.) Number 3. (OK.) Number 4. (OK.) Number 5. (Five second interval.) Very good. This is a task in auditory perception. The subject is to count the number of clicks of the metronome. The clicks of the metronome are recorded in order that I may be sure of the number of clicks presented each time. Since I will be recording the responses of all five subjects from the control room, it will be easiest for me if each of you report your answer in a given order. Please report your answers according to the number you were given when you entered the experimental room. That is, the subject having number 1 will report his answer first, the subject having number 2 will report second, and so on. Please remain silent, except for reporting your answers, during the experiment. Are there any questions? (Questions were answered by rereading appropriate parts of the instructions.) All right, here we go. Listen carefully and tell me the number of clicks when it is your turn.

At this point, 20 series of clicks were presented in the order shown under Pretest in Table 1, and the four background voices reported as indicated in the table, after each presentation. Clicks range from 22 to 26 in each

TABLE 1
NUMBER OF METRONOME CLICKS AND NUMBER REPORTED BY EACH BACKGROUND VOICE

Pretest		Reinforcement period	Posttest	
Number of clicks	Number reported		Number of clicks	Number reported
25	24-24-24-24	23*	25	24-24-24-24
26	26-26-26-26	26	26	26-26-26-26
24	25-25-25-25	22*	24	25-25-25-25
25	26-26-26-26	24	25	26-26-26-26
23	24-25-25-24	23*	23	24-25-25-24
24	23-23-23-23	26	24	23-23-23-23
23	22-22-22-22	25*	23	22-22-22-22
22	23-23-23-23	22*	22	23-23-23-23
25	24-23-24-23	24	25	24-23-24-23
26	27-27-27-27	22*	26	27-27-27-27
22	21-21-21-21	26	22	21-21-21-21
24	24-24-24-24	25*	24	24-24-24-24
22	21-21-21-21	22	22	21-21-21-21
25	26-26-26-26	23	25	26-26-26-26
26	27-27-27-27	24*	26	27-27-27-27
23	22-22-22-22	23*	23	22-22-22-22
24	23-23-23-23	25	24	23-23-23-23
22	22-22-22-22	24*	22	22-22-22-22
26	25-25-25-25	26	26	25-25-25-25
23	24-24-24-24	25	23	24-24-24-24

* Signifies the responses to be reinforced or punished on the 50 per cent PR and PP schedules.

series, since Blake, Helson, and Mouton (5) found clicks in this range to be effective for this task. The metronome was set at 200 beats per minute, and the background voices differed from the actual number of beats presented by one beat in 15 trials because Blake *et al.* (5) found these conditions to produce the greatest amount of conformity behavior in their study. Each number of clicks was presented four times with different background voice conditions. The order in which the differing number of beats was presented was randomly determined.

As seen in Table 1, in eight out of the 20 presentations, all four background voices reported one less click than the actual number presented; in seven presentations, the four voices reported one more click; in one presentation, two "subjects" reported one more click and two "subjects" reported two more clicks; in one presentation, two "subjects" reported one less click and two "subjects" reported two less clicks; and, in three presentations, all four background voices reported the number of clicks actually presented.

The three presentations in which the background voices reported correctly and the two presentations in which two background voices disagreed with the other two were included in an effort to minimize the suspicion of the subjects in the reality of the situation.

Blake *et al.* (5) had their background subjects report the correct number of clicks in three out of nine trials in an effort to keep the subjects' faith in the task. In addition to this, in the present study, two background voices differed from the rest in two trials because of the additional 11 trials in the present study. It was assumed that if the background voices disagreed in two of the trials, the subjects would be more convinced of the reality of the situation. The taped background voices responded at intervals ranging from one to three seconds, and the successive stimulus presentations followed the response of the fourth background voice by five seconds. The four background voices responded in the same order each time.

After the final stimulus presentation in the Pretest period, the experimenter turned off the tape recorder and waited five minutes before entering the experimental room. He then said the following:

> Before we continue with Series 2, let's try it a number of times individually for practice. I will now turn on just the record with the clicks. None of the other subjects will be able to hear us. Ready now, listen carefully and tell me the number of clicks.

At this point, the experimenter left the room, turned on the tape recorder, and returned to the experimental room. Twenty presentations of clicks were given in the order shown in Table 1 under the heading Reinforcement Period (simply the reverse order in which they were presented in the Pretest), and subjects were treated in one of five ways. After every response of each member of the Continuous Reinforcement Group (CR), the experimenter said "right." Following the responses of each member of the 50 per cent Partial Reinforcement Group (PR), the experimenter said "right" after 10 responses and said nothing after the other 10 responses. The responses to be reinforced were previously determined in a random fashion. After each response of the members of the Continuous Punishment Group (CP), the experimenter said "wrong." Following the responses of each member of the 50 per cent Partial Punishment Group (PP), the experimenter said "wrong" after 10 randomly determined responses, and said nothing after the other 10 responses. And, after the responses of each member of the Control Group, the experimenter said nothing.

After the subjects had completed one of these treatments, they were given the following instructions:

> All right, now we will continue with Series 2 in the same manner that we did Series 1. Ready now, listen carefully and tell me the number of clicks immediately when it is your turn.

The experimenter then left the experimental room and went into the adjoining observation room. The stimulus tape then began once more with the experimenter's voice saying: "All right, here we go with Series 2." Twenty presentations of clicks were then presented in the order shown under Posttest in Table 1, and the four background voices reported as indicated in the table after each presentation.

Following the Posttest, all subjects were asked to explain what they believed to be the object of the experiment. Subjects stating disbelief about the presence of four other subjects in adjoining rooms were eliminated from the study.

The dependent measure in this study was the number of reported clicks by which subjects differed from the background voices on each stimulus presentation. In the present study, this number is referred to as a subject's deviation score. If, for example, on one trial the four background voices reported 24 clicks and the subject reported 27, his deviation score is 3. If on another trial the background voices reported 24 and the subject reported 22, his deviation score is 2. Deviation scores were obtained for each subject on each of the Pretest and Posttest trials where the background voices unanimously differed from the actual number of clicks presented. Deviation scores were used because they indicate precisely the extent of each subject's conformity or lack of conformity on every response.

C. RESULTS

A factorial analysis of variance with repeated measures was used to analyze the experimental data. The means of all groups are shown in Table 2. The results of the analysis of variance are summarized in Table 3. This table indicates that there is a significant difference between groups when Pretest and Posttest scores are summed and the treatment groups are compared with one another. At the .05 level an *F* value equal to or greater than 2.50 is necessary for significance, and an *F* of 2.568 was obtained.

However, since the major interest of this study was to examine the effects of the different conditions, measured over the Pretest and Posttests, the interaction between these variables assumes major importance in the results. As seen in Table 3, the effect of this interaction was significant well beyond the .05 level. At the .05 level an *F* value equal to or greater than 2.50 is necessary for significance, and an *F* of 8.967 was obtained. The Multiple Range Tests indicated that all Pretest and Posttest group means (indicated in Table 2) significantly differ at the .05 level for all groups except the Control Group.

Table 3 also indicates that the total of the Pretest scores for all groups did not significantly differ from the total of the Posttest scores at the .05 level.

TABLE 2
Mean Number of Deviation Scores Per Subject over Fifteen Stimulus Presentations

Condition	Pretest	Posttest	Difference
Continuous Reinforcement	11.00	13.60	2.60
Partial Reinforcement	11.73	15.00	3.27
Continuous Punishment	11.00	9.47	—1.53
Partial Punishment	10.67	9.00	—1.67
Control	11.27	11.13	-0.14

TABLE 3
Analysis of Variance of Deviation Scores, Condition by Pretest and Posttest

Source	*df*	Sums of squares	Mean squares	*F* ratio
Condition (C)	4	255.97	63.99	2.56*
Error between	70	1744.60	24.92	
Pretest and Posttest (T)	1	9.63	9.63	2.162
C by T	4	159.71	39.93	8.967**
Error within	70	311.67	4.45	
Total variation	149	2481.57		

* *p* less than .05.
** *p* less than .01.

It must be stressed that, although the general effects of the punishing and reinforcing conditions differed considerably from one another, the effects of continuous and partial reinforcement were very similar; and the same holds true for the effects of continuous and partial punishment.

As seen in Table 2 partial reinforcement and continuous reinforcement resulted in a decrease in conforming responses, with partial reinforcement having a slightly greater effect in this direction. Partial punishment and continuous punishment resulted in an increase in conforming responses, with partial punishment having a slightly greater effect. The Control Group Subjects deviated from the background voices by about the same amount in the Pretest and Posttest.

Table 4 indicates the percentages of conforming responses in the Pretest and in the Posttest. While in the Pretest, the subjects in the Partial Reinforcement and Partial Punishment Groups conformed on approximately the same percentage of responses (the former 33.8 and the latter 32.0); after the Reinforcement Period these percentages were markedly altered (Partial Reinforcement Group, 15.5 and Partial Punishment Group, 41.8). These treatments produced the greatest effects.

For total performance, no significant differences were found between the Pretest and the Posttest, but when this variable was analyzed by conditions a significant difference was indicated. The differences, therefore, appear to result from the conditions employed.

D. Discussion

The most significant findings derived from this study, in the writer's opinion, were the differing effects resulting after the reinforcement and the punishment schedules.

Following the schedules of reinforcement and punishment, subjects behaved in a very similar manner to subjects responding to other tasks under the same schedules using different reinforcers and punishers (13, 19). That is, reinforcing conditions produced an increase in the rewarded response class after those conditions were eliminated, and punishment resulted in a decrease in these responses after the punishing conditions were eliminated. And, finally, subjects in the control group responded in approximately the same manner in the Posttest as in the Pretest. The results of this experiment indicate, therefore, that different reinforcement histories (schedules of reinforcement) following an individual's responses to a task can affect his "conformity" behavior with regard to that task when in a simulated small group situation. This further suggests that "conformity behavior" as investigated in this study is a workable response class on which the effects of verbal reinforcement can be studied.

In the Pretest, subjects in all five groups responded very similarly. In this phase of the experiment subjects conformed on approximately the same percentage of responses that Blake (5) reported on a highly similar task (about 33.3 per cent). Also as Blake (5) and Asch (2) reported, subjects ranged from almost complete independence of group opinion to nearly complete submission in the Pretest. After the Reinforcement Period, however, these percentages were markedly altered (see Table 4). Even the reports of several subjects whose responses were extremely nonconforming and of some subjects whose responses were highly conforming changed markedly in the opposite direction. This would suggest, perhaps, that at least part of the variability in conforming behavior found by Asch (2) and Blake (5) and found in the Pretest of the present study was a result of the behavioral histories of the subjects in related tasks or in tasks in general. It would, therefore, seem profitable to investigate further the effects of built-in behavioral histories on subjects in tasks such as this if these differences in the "conformity" behavior of subjects are to be understood. Perhaps, such research would make it possible to predict and control this behavior for each individual, in addition to simply measuring it and reporting an average trend. The results of this study would suggest that highly conforming subjects in tasks of this kind possibly have received a lot of "punishment" in similar or related tasks, while nonconforming subjects might have been reinforced a good deal.

TABLE 4

PERCENTAGES OF CONFORMING RESPONSES IN CONDITIONS AND IN PRE- AND POSTTESTS

	CR	PR	CP	PP	Control	Total
Pretest	33.8*	33.8*	30.7	32.0*	26.7	31.4
Posttest	18.2	15.5	37.3	41.8	28.4	28.3

* Indicates a significant difference in Pre- and Posttest percentage at the .05 level.

Although, for the most part, subjects responded to the experimental treatments in the hypothesized manner, some subjects in each group responded differently. The reasons for this within-group variability are unclear. For example, two subjects conformed more after reinforcement and some conformed less after punishment. Perhaps part of the variability was related to the ambiguity of the meaning of the experimenter's silences. While some subjects might have responded to the experimenter's silences as meaning they were wrong, others might have responded to the silences as meaning they were right, while still others might have responded to the silent experimenter as simply a recorder of their responses. Another possible source of variability concerns the reinforcing values of the experimenter's "right" and the punishing value of "wrong" for different subjects. The effects of these words might be quite different among subjects due to their various backgrounds. Still another source of variability was the awareness level of the subjects. While some subjects might have been convinced of the presence of other legitimate subjects in adjoining rooms, others might have gradually become more suspicious as the experiment progressed, and as a result conformed less. Further research might well be done to investigate this within-group variability and to measure the effects of other schedules of reinforcement on "conforming" behavior.

In general, the results of this experiment suggest that the "conforming" behavior of each individual in a group (or, at least, in a simulated group), like other behavior, is related to previous learning and that, given the proper conditions, this behavior might be predicted and controlled in the same manner as other behavior of the individual. In other words, it is suggested that the "group behavior" of each individual in a group might be dealt with in very much the same manner as any other aspect of his behavior and that, if the relevant variables were studied, this behavior might be found to operate under the same principles or laws as other behavior.

E. Summary

The present study investigates the effects of four verbal reinforcement schedules—continuous reinforcement, partial reinforcement, continuous punishment, and partial punishment—on the responses of subjects to a task in a simulated group situation.

Seventy-five subjects were dealt with. These were subdivided equally over the four treatment groups and the Control Group.

The experimental task required subjects to count the number of clicks of a metronome and to report that number after four "other subjects in adjoining rooms" had reported. Actually, the four voices heard by the subjects were the recorded voices of graduate students reporting incorrectly on most trials. Twenty series of clicks were presented twice with the experimental conditions being introduced between presentations. The dependent measure consisted of the number of reported clicks by which a subject deviated from the false "subjects."

The data were analyzed by an analysis of variance with repeated measures and by a series of Duncan Multiple Range Tests.

Differences in "conforming" behavior following the four reinforcement and punishment schedules were found to be significant beyond the .01 level

when analyzed by Pretests and Posttests. Reinforcing experimental treatments resulted in an increase in nonconforming responses, while punishing treatments resulted in a decrease in this response class.

It was concluded that verbal conditioning methodology appears to offer a road to the study of certain aspects of "group" behavior.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgment. In H. Guetskow (Ed.), *Groups, Leadership and Men.* Pittsburgh: Carnegie, 1951.
2. ———. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monog.*, 1956, **70**, No. 9 (Whole No. 416).
3. BACHRACH, A. J., CANDLAND, D. K., & GIBSON, J. T. Experiments in verbal behavior. I. Group reinforcement of individual response. In I. Berg & B. Bass (Eds.), *Conformity and Deviation.* New York: Harper, 1961. Pp. 258-285.
4. BLAKE, R. R., & BREHM, J. W. The use of tape recording to simulate a group atmosphere. *J. Abn. & Soc. Psychol.*, 1954, **49**, 311-313.
5. BLAKE, R. R., HELSON, H., & MOUTON, J. S. The generality of conformity behavior as a function of factual anchorage, difficulty of task, and amount of social pressure. *J. Personal.*, 1957, **25**, 294-305.
6. BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. Exper. Psychol.*, 1956, **52**, 283-287.
7. FATTU, N., & MECH, E. V. Resistance to extinction of verbal responses following two patterns of reinforcement. *J. Genet. Psychol.*, 1955, **53**, 193-198.
8. GREENSPOON, J. The effect of verbal and nonverbal stimuli on the frequency of members of two verbal response classes. Unpublished Ph.D. dissertation, Indiana University, Bloomington, 1951.
9. ———. The effect of two nonverbal stimuli on the frequency of members of two verbal response classes. Paper presented at the American Psychological Association Convention, New York City, 1954.
10. ———. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
11. ———. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), *Experimental Foundations of Clinical Psychology.* New York: Basic Books, 1962.
12. GROSSBERG, J. H. The effect of reinforcement schedule and response class on verbal conditioning. Unpublished Ph.D. dissertation, Indiana University, Bloomington, 1956.
13. KANFER, F. The effect of partial reinforcement on acquisition and extinction of a class of verbal responses. *J. Exper. Psychol.*, 1954, **48**, 424-432.
14. ———. Verbal conditioning: Reinforcement schedules and experimenter influence. *Psychol. Rep.*, 1958, **4**, 443-452.
15. KIRMAN, W. J. The relationship between learning, with and without awareness, to personality needs. Unpublished Ph.D. dissertation, Columbia University, New York City, 1958.
16. KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **15**, 148-171.
17. MCNAIR, D. M. Reinforcement of verbal behavior. *J. Exper. Psychol.*, 1957, **53**, 40-46.

18. Salzinger, K. Experimental manipulation of verbal behavior: A review. *J. Genet. Psychol.*, 1959, **61**, 65-95.
19. Sandler, J. The effect of negative verbal cues upon verbal behavior. Unpublished Ph.D. dissertation, Florida State University, Tallahassee, 1958.
20. Solley, C. M., & Santos, J. F. Perceptual learning with partial verbal reinforcement. *Percept. & Motor Skills*, 1958, **8**, 183-193.
21. Timmons, E. C. Experiments in conditioning operant verbal behavior. Unpublished Ph.D. dissertation, University of Tennessee, Knoxville, 1959.

Department of Behavioral Science
Institute of Physical Medicine and Rehabilitation
400 East 34th Street
New York, New York 10016

Published as a separate and in *The Journal of Psychology*, 1966, 62, 111-137.

GROUP EFFECTIVENESS IN CREATIVE PROBLEM-SOLVING TASKS: AN EXAMINATION OF RELEVANT VARIABLES*[1]

Department of Psychology, University of Arizona

SALVATORE V. ZAGONA, JOE E. WILLIS, AND WILLIAM J. MACKINNON

A. INTRODUCTION

In this age of increased specialization and rapid expansion of knowledge, there is an increasing tendency to rely on teams of experts for the solution of problems. This is especially true concerning research and development areas in industry and government. In these cases, the creativeness of a group may have effects which could—directly or indirectly—influence immense fortunes and the welfare of nations.

Presumably any condition tending to shape the structure or function of a group would also, in some way, contribute to influence the manner in which the group solves a variety of problems facing it. In the past few decades, research in group dynamics has contributed much in theory and method in this area. However, the purpose of the present paper is to investigate group process only as it relates to effective creativity in problem-solving tasks. The emphasis will be upon creativity as a psychological dimension of growing research importance. This variable will be investigated within the context of group interactions; only group processes that bear directly upon the expression of creativity by groups will be considered.

The concept "creativity" is not easily defined, but since this paper is concerned with problems which require creative solutions, a consideration of the concept in depth is necessary.

The literature of psychology today reveals the term "creative" used as an adjective to modify such diverse things as "ability" (14, 47), "process" (14, 39), "behavior" (13, 15), "thinking" (15, 39), "product" (35, 47), and so on. All these adjectival uses have, in turn, been employed to describe the following: personality, people, thoughts, behavior, and products, in such diverse fields as mathematics, physics, music, social science, painting, and poetry—to name but a few. Yet few people seem to find it inconsistent or

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 13, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The authors wish to express their appreciation to Dr. Jack E. Capehart for his helpful suggestions in the course of preparation of this manuscript.

peculiar to talk about "creativity" as though it were a *single* variable that should be isolated and studied.

This state of affairs with respect to the status of "creativity" is rather similar to the tale of the blind men trying to describe the shape of an elephant of which each blind man was touching a different and separate part. One described the trunk, another a leg, and a third the tail, but each insisted that his was *the* description of the elephant.

The progress of the concept of creativity to its present research status, although it has a less compelling connotation, is analogous in some ways to the concept "disease," in which distinct diagnostic entities were isolated and investigated only through the differentiation and classification of recurring characteristic symptoms. It would seem that the same sort of differentiation must be made with regard to the "creativity" concept.

Flanagan (11) has distinguished among "productivity," "creativity," and "ingenuity." "Productivity," he said, relates to the offering of many new ideas, and emphasizes both quantity and contribution. "Creativity" consists in bringing something new into being, with emphasis on the lack of previous existence of the product or idea. "Ingenuity" denotes inventing or discovering a solution to a problem, with some emphasis on unusualness or "surprise" in its manner of expression. His definition of "ingenuity" corresponds in large part to the definition of "creativity" implicit in most of the studies here reviewed.

Sprecher (46, p. 77) has recognized this semantic problem and has proposed an approach for its solution. "Since human beings assign meaning to the term, they may well choose not to agree; it is not necessary that they should always agree, but only that they should know how and to what extent they disagree." His position is that if we can get people to state as precisely as possible what they mean by "creativity," then any disagreement which is discovered will be tolerable.

Haefle (14) defined "creativity" as the ability to make new combinations between two or more concepts already in the mind. Schutz[2] used the term "creative behavior" to describe behavior that is unusual, satisfying, or effective. He also related the concept "divergent thinking," used by Guilford and others, to the definition of creativity. "Divergent thinking" is applied to situations which require the use of multiple solutions rather than problems with just one correct answer. This latter aspect is a necessary condition for group creative problem solving.

---

[2] Schutz, W. C. Creative behavior: Training and theory development. Personal communication, 1963.

Fiedler (7, p. 308) has provided a definition of group creative problem solving that would seem to be adequate for the purposes of this paper. He defined it as ". . . group interaction leading to an original solution of a problem or to a more effective recombining of given elements into new forms."

There are some [e.g., Flanagan (11)] who maintain that the creative product must have never before existed. This seems to be a valid position, particularly since he substitutes a different concept (ingenuity) to denote the specific steps or operations of creative problem solving. The many other studies of creative problem solving can be viewed as meeting the criteria for creativity by stipulating that the "creative" idea or product must occur for the first time in the life of the person concerned. Actually, whether or not the same solution has been produced before by someone else is not of crucial importance. The essential point is that it must be a first occurrence within the lifespan of the person concerned. If we could not make this stipulation, there could be no basis for experimental studies of creativity, since one of the most important characteristics of a scientific investigation is that it be subject to replication.

The material presented in the present paper will be organized in relation to five basic variables which have been investigated in past research on group creative problem solving: (*a*) task variables, (*b*) individual *vs.* group comparison, (*c*) the role of leadership, (*d*) the influence of training, and (*e*) how group characteristics such as structure and function relate to its creative expression.

B. Task Variables: The Problems

A convenient method of classifying problems requiring creative solutions is that used by Krech and Crutchfield (22). They divided the problems into three categories: *Explanation*, *Prediction*, and *Invention*, each of which is distinguished by the different goal it involves. The goal of "Explanation" is to understand why a certain event has occurred. The products of people like Galileo, Einstein, and Newton would fall in this category. For "Prediction" the goal is to understand future consequences, given certain conditions. The goal of "Invention" is to create a set of new conditions which will result in a specified event.

These three classes of solutions are also distinguished in terms of the degree to which the starting stimulus pattern is structured. In the "Explanation" category the stimulus pattern which confronts the problem-solver is, to a large degree, specified. The specific objects involved have been observed previously to do specific things in specific sequences and proximities. At least in some cases the stimulus pattern consists of objects which are concretely defined, as are their interactions. These problems readily yield a "correct" perception of the event even though it may be a novel one.

In "Prediction" problems, the probability of having the correct organization of the stimulus condition "forced"—i.e., brought about automatically—is reduced because the objects and the relations between them are fewer and not so well defined as in "Explanation" problems. Only the antecedent conditions are known, and the event which will occur in the future must be predicted on the basis of them.

The final category, "Invention," is the most difficult type of problem in terms of stimulus pattern. The "Invention" problem has fewer clues to solutions in the stimulus pattern and accordingly makes greater demands on the resources of the individual. The inventor's problem is to bring about an event that is specified before he begins, and that specified event does not and never before has existed. He must not only attend to the givens, as in a "Prediction" problem, but he must also be aware of an extensive number of objects and relationships beyond the given stimulus pattern. The inventor's is the least structured of the three types of stimulus patterns.

Let us now turn to the various problems used in experiments on group creative problem solving. The problems will be classified in terms of the three categories above, but these classifications reflect only the views of the present authors, not necessarily the views of those who originally conceptualized and carried out the studies cited. Among these studies more than 24 separate problems were used which required a creative solution of some kind. Some overlapping in the classification is inevitable. Space does not permit a full description of all of the problems, but enough to be representative of trends will be presented, along with references.

1. *Explanation*

a. "Moral Justification" problem. Fiedler (7) asked an experimental group composed of church administrators and lay leaders to write a 250-word statement in which they were to justify a decision by the minister of their congregation not to divulge the intention of a local physician to commit a mercy killing.

b. "Tower of Hanoi" problem. The group is confronted with three pegs, one of which has three rings of differing sizes on it. The problem is to explain how to transfer all of the rings to another peg without pulling off more than one at a time and without putting a large ring on top of a smaller one (23, 38).

c. "Tartaglia Transportation" problem. Three jealous husbands and their beautiful wives must cross a river in a boat which will hold only three persons at a time, with the constraint that no husband will allow his wife to be in the presence of another man unless he is also present. Only husbands can row (23, 24, 25).

2. *Prediction*

a. "Thumbs" problem. What would be the consequences if everyone born after 1960 had an extra thumb on each hand? (5, 6, 40).

b. "People" problem. What would be the consequences if over a period of 20 years the average height of people increased 80 inches and the average weight doubled? (6).

c. "Discharge" problem. This problem was developed by Cohen *et al.* (5) especially for a group composed of professional and administrative personnel in a psychiatric hospital. The problem concerns the practical benefits and difficulties which would ensue if all patients were able to return to their families.

3. *Invention*

a. "Mined-Road" problem. Originally developed by the OSS Assessment Staff during World War II, this problem has been used in many of the studies reviewed. It requires that the subject group formulate a plan of action for getting a group of five men across a road which is mined with supersensitive enemy mines which cannot be dug up or neutralized. The road is 12-feet wide and bordered on both sides by trees about 40-feet tall. Scattered around on the approach side of the road are ropes, beams, a discarded auto tire, pulley, and a variety of other potentially usable material and debris. The groups write about their solutions to the problem and the solutions are then analyzed by a "Quality Point System" (QPS), which is a standardized method for evaluating different degrees of quality in solutions (16, 18, 19, 26, 27).

b. "Change of Work Procedures" problem. This and the following three problems are role-playing problems. In this one, one of the members is randomly assigned the role of a foreman and the other three members the role of workers. The problem goes like this: An efficiency expert has found that the method of work the group has been using is inefficient—i.e., a lot of time is being wasted. The group members had been alternating jobs in assembling carburetors, so that each worked an equal amount of time on each of the three phases. The group was required to decide whether (*a*) to adopt a new procedure in which each stays on the same phase all day (thereby increasing production, but also boredom); (*b*) to maintain the *status quo;* or (*c*) conceiving the task as a creative problem, to search for an "integrative" solution which would combine the best features of the *status quo* and new procedure so as to satisfy both workers and management. This integrative solution is considered to be a creative one (16, 19, 20, 29).

c. "Student Assistance" problem. There are two principal parts to this problem. Each of the four participants is assigned the role of a particular type of student who is eligible for one reason or another to receive some university funds. The first task, of allocating the funds among themselves, was designed to create conflict in the group. The second task requires the participants to suggest a method for determining the people who should receive money in the future (19).

d. "Painter-Inspector Argument" problem. The four roles are a painter, an inspector and rough primer, a shop steward, and a foreman. They meet to settle a conflict between the inspector and the painter. The argument reflects a basic problem which exists in the entire shop, but apparently few groups are capable of getting beyond the superficial problem to identify the total shop problem (18).

e. "Point Distribution" problem. For this problem the group members, who are students, divide among themselves a certain number of points, which were assigned to each group. These points were to be added to the total number of points which each student accumulated over his four class examinations and would thus help determine his final grade (19).

We return now to non-role-playing problems, in which no roles are assigned to the members of the group.

f. "Tourist" problem. The members of the group must devise ways of getting more Europeans to come to this country during their vacations (5, 6, 40, 49).

g. "Education" problem. What can be done to cope with the increasing number of children of school age? (2, 6, 40).

h. "Library" problem. A library has been plagued with the mutilation of magazines and books, with the cutting out of pages and the stealing of books. What can be done? (2).

i. "Hanger" problem and "Broom" problem. The members of the group are asked to list as many uses as they can for these items (35, 43).

j. "Bus" problem. How can a shy girl go about meeting a boy she regularly sees on a bus? (40).

k. "Argument Construction" task. The *S*s must develop arguments for and against the use of a rigorous and dangerous military program which could be expected to result in a high casualty rate during training, but which also would lead to a low death rate during actual battle conditions (1).

l. "Fame and Immortality" task. How can a person of average ability achieve fame and immortality even though he does not possess any particular talents? (1, 51, 52).

m. "Church" problem. The *S*s are to write solutions to the following problem: A church which has completed nearly two-thirds of its building ran out of money. Since it is in an area which has been blacklisted by banks and loan associations because it is a dying area, the church cannot obtain credit. How can it obtain the money needed to complete the building? (51, 52).

Fiedler (7) used a problem in which groups were given a TAT card and instructed to devise three stories about it within 20 minutes. These stories were to be as different and original as possible. They were rated by two independent judges, using a scoring manual. He and his co-workers also used the following three problems with a group composed of church administrators and lay members: (*a*) invent a Sunday School parable for six- to eight-year-olds which would illustrate the desirability of the doctrine of separation of church and state, (*b*) develop an appeal for funds to send a ministerial student through divinity school, and (*c*) plan a three-minute skit to illustrate the need for better music in worship service (9).

Triandis, Bass, Ewen, and Mikesell (52) presented their experimental groups with the task of suggesting new legislation which was inventive, original, and creative enough to solve the problem of implementing the Kennedy programs of The Peace Corps, Federal Aid to Education, School Desegregation, etc. Each solution was mimeographed and Thurstone's successive-intervals procedure was used to obtain indices of creativity on an equal-interval scale.

For another task used by Triandis *et al.* (52) the groups were asked to write an original creative story which contained the five concepts: "Evolution," "Army," "Federal Aid to Education," "Labor Unions," and "Immortality."

Other problems which have been used in research on group creative problem solving are (*a*) the Guilford "Unusual Uses" Test (34); (*b*) the "Plot Titles" Test (34); (*c*) the invention of brand names for a cigar, a deodorant, and an automobile produced especially for Purdue students (54); and (*d*) writing as many captions as possible for a Saturday Evening Post cartoon (56).

C. TASK VARIABLES: PROBLEM EFFECTS

A number of studies have discovered complications and areas requiring caution with regard to the tasks used in experiments in group creative problem solving. This section will bring these important findings together so that they may be examined in closer relation to each other.

Lorge *et al.* (26) presented to groups the "Mined-Road" problem in four degrees of remoteness from reality: (*a*) a verbal description; (*b*) a photographic representation; (*c*) a miniature scale model, but without permitting manipulation of materials; and (*d*) the miniature scale model, allowing the manipulation of the materials. Results indicated that there were no significant differences among the solutions for the four levels of remoteness from reality.

Using the "Mined-Road" problem constructed in reality, Lorge *et al.* (27) also compared the actual accomplishments of groups (as recorded by trained observers) with reports written by teams and by individuals. Lorge *et al.* found that the teams wrote a solution which gave a scoring value somewhat inferior to their actual performance on three of the four major components used in analyzing the quality of the solutions. The opposite was true of the individuals.

Parloff and Handlon (40) reported two items of interest with regard to task variables. They made a distinction between "generated" ideas and written ideas. "Generated" ideas referred to those ideas which came up in group discussion but which were not recorded. A distinction was also made between "real" and "unreal" problems. Included among the "real" problems were the "Education" problem, the "Tourist" problem, and the "Bus" problem. The "Thumb" problem is an example of an "unreal" problem. They found that more ideas were unreported for unreal than for real problems.

Hoffman (16) made a distinction between problems which involve quality alone (e.g., the "Mined-Road" problem) and problems which involve quality and acceptance by the group (e.g., the "Change of Work Procedures" problem).

Hoffman and Smith (19) found the differences among the problems they used to be a major source of variance. The four problems they used were (*a*) the "Mined-Road" problem, (*b*) the "Student Assistance Fund," (*c*) the "Change of Work Procedures," and (*d*) the "Point Distribution" problems. They suggested that it would be useful to classify problems in terms of the dichotomies: quality *vs.* acceptance and creativity *vs.* decision-making.

When groups were asked to solve the same problem for a second time, it was found that a significantly higher percentage of "integrative" (creative) solutions resulted (28). It was hypothesized that by satisfying the leader's needs to obtain a solution and then returning again to the problem, the groups became more problem-minded and the leader was more willing to entertain the viewpoints of the other group members.

Maier and Solem (31) found that they could increase the number of alternatives which came up for consideration and also the number of good creative solutions by the use of instructions calculated to turn choice situations into problem situations.

Meadow, Parnes, and Reese (35) found that significantly more good solutions were produced when brainstorming instructions were given first than when prior training did not include brainstorming instructions. Parnes (41) reported a study which showed that significantly more good ideas appeared in the final third of a subject's list of ideas than in the first two-thirds of the list. This provides an indication that extended effort on a problem which requires creative thinking can be rewarding in terms of an increased proportion of good ideas among those produced in the extra time spent on the problem.

Lorge and Solomon (24, 25) warn us that previous experience with the same or a related problem by members of the group being studied can cause considerable trouble, especially in experiments with "Eureka-type" problems ("Tartaglia," "Tower of Hanoi"). They found that reports by *S*s as to previous experience with problems are unreliable.

Anderson and Fiedler (1) studied the effect of two types of leader-member interaction using four separate creativity tasks: "TAT stories," "Unusual Uses" test, "Argument Construction," and "Fame and Immortality." The results showed a strong relationship between the type of task, leadership form, and the leader's individual attributes. In fact, most of the results were task specific. This study especially emphasizes the importance of identifying carefully the nature of the creative task used in experiments.

Hoffman (17) has suggested some theoretical considerations which might facilitate studies of group creative problem solving. He stipulated four conditions which should be present in order that a creative solution to a problem might emerge. First, differing (but comparable) cognitions must coexist. In a group situation, satisfaction of this condition would involve at least two contributions concerning a problem, ordinarily from different group members. Such contributions normally take the form of possible solutions. At this point we must introduce the concept of valence. Valence is the degree of acceptability a cognition may have for a problem-solving unit (individual or group). For groups it is translated to mean the degree to which a group member's suggestion is acceptable to all members of that group.

The second condition is that at least two differing cognitions must acquire greater positive valence than the minimum threshold value and must be approximately equal in value, so that none of the alternatives can be accepted and an impasse is reached. A creative solution to such a problem situation would involve exploiting the advantages of both alternatives and possibly would incorporate, with advantage, seemingly incompatible facts which support the alternatives.

The third condition is that the problem solving must occur in a situation in which the problem-solving unit is required to arrive at the best possible decision. (This may be considered the boundary condition.) Hoffman considered it likely that failure to satisfy this third condition has probably been the source of many people's failure to solve problems adequately. The fourth condition is that the cognitive components associated with or contributing to the valence of each of the conflicting alternatives should be abstracted and their points of conflict recognized.

D. INDIVIDUAL *vs.* GROUP-EFFORT

In studies concerned with group creativity, the fundamental question arises: "Are two (or more) heads better than one in formulating creative solutions to problems?" Since creative ideas can originally come only from individuals, the question is whether or not individuals can be more productive in groups than working as individuals. To answer this question, studies have been conducted that compare individuals and groups with respect to results obtained by each, and with respect to factors employed by each in arriving at the final solution. Different studies have provided data that in some instances support the group's superiority over individuals and in others the individual's superiority over groups.

Shaw (44) compared the ability of individuals with that of cooperative groups of four persons in solving complex problems. Her results appeared to indicate group superiority in this area due, in part, to the group's larger proportion of correct solutions achieved through the rejection of incorrect suggestions, and checking of errors. In an extended replication of this study, Marquart (32) combined individuals into nominal groups of three because Shaw's method of analyzing the data did not indicate whether the group solution was due only to the best member of the group. Marquart, using her method, found that the groups, working as groups, were no better than groups working as individuals.

Taylor and Faust (48), using a variation of the "Twenty Questions" problem, found the performance of groups to be superior to that of individuals on the dependent variable measures: number of failures to reach solution, number of questions asked, and elapsed time per problem. In terms of man-minutes required to reach solution, however, performance of individuals was superior to that of groups.

Taylor, Berry, and Block (49) designed an experiment to determine whether group participation when using brainstorming facilitates or inhibits creative thinking. Ninety-six *S*s were divided into 12 groups of four, and a pool of 48 individuals. After the groups and the individuals had independently brainstormed solutions to the "Tourist," "Thumbs," and "Education" problems, the solutions written by individuals were randomly divided into 12 "nominal" groups of four. The achievements of these nominal groups were then a measure of the performance that could be expected in the absence of an inhibitive or a facilitative effect of group participation. Their results indicated that the "nominal" groups were superior to real groups with respect to (*a*) the mean total number of ideas, (*b*) the mean number of unique ideas, and (*c*) three measures that took into consideration the proportion of ideas produced to quality of the ideas. Taylor *et al.* concluded that group brainstorming inhibits creative thinking.

A study by Dunnette, Campbell, and Jaastad (6) replicated the study just described, but extended it by exposing both groups (nominal and real) to both conditions. That is, *S*s were required to solve two sets of problems, one set under the group brainstorming condition and one set under brainstorming as individuals. The population was also extended by using research scientists and advertising personnel from an industrial firm. Results were similar to those of Taylor *et al.*, and Dunnette *et al.* also concluded that group participation, even under brainstorming conditions, has an inhibitory influence.

According to Osborn (39), the basic rules of brainstorming are (*a*) positively no criticism is allowed; (*b*) "free-wheeling" is encouraged; (*c*) quantity of ideas is desirable; and (*d*) combination and improvement of ideas are encouraged. Osborn further suggested that the optimal group size is 12 (p. 169), and that at least the group discussion leader should have extensive training in brainstorming. It is quite possible that either of these variables could have accounted for the resultant superiority of individuals over groups in the studies by Taylor *et al.* (49) and by Dunnette *et al.* (6). Both of these studies used groups of four, and neither trained *S*s in the brainstorming technique for more than an hour total.

Parloff and Handlon (40) tested the hypothesis that techniques which reduce critical judgment do not necessarily increase the generation of creative ideas, but instead increase *S*'s willingness to report more of the ideas which ordinarily he would dismiss as unworthy. To effect this comparison Parloff and Handlon used dyads, and tape-recorded all spoken ideas (which they called "generated" ideas), and also had *S*s write their ideas. Two types of instructions were used: under the "low critical" instructions, the group members were to avoid criticism at all costs and were to give their imaginations free rein; under the "high critical" instructions, they were instructed to be analytic so that only the "good" idea would be selected. There were significantly more written solutions and more "good" written solutions by the dyads under the low critical condition than under the high critical condition. However, it was also found that the groups under the high critical condition "generated" more ideas than did the low critical groups.

These results were interpreted to mean that the suspension of critical judgment simply lowers *S*s' standards for reporting ideas, without substantially increasing their repertoire of good ideas. What seems to be more likely is that the low-critical dyads simply could not be "generating" ideas while they were writing them. In other words, while the low-critical dyads were writing ideas, the high-critical dyads were "spinning wheels"—that is, talking without making serious attempts at finding solutions.

Lorge and Solomon (23) proposed two models of group problem-solving behavior for the Eureka-type problem: (*a*) the probability of a group solution is a function of the probability that the group will include at least one person who can solve the problem, and (*b*) group solutions depend upon the pooled abilities of the group members. Lorge and Solomon showed that Shaw's data (44) could be explained as an outcome of the type of behaviors suggested by the models, but warned that that the models may not be appropriate with special populations and special kinds of problems.

In a later study Lorge and Solomon (24) found that previous experience with the problem ("Tartaglia Transportation" problem) or its solution increased the probability of solution by individuals and by groups. (This was in accord with their model.) They concluded from this that any previous experience with a problem should be included in the interpretation of results. This observation was confirmed again in a still later study (25).

Recall the study by Lorge, Tuckman, Aikman, Spiegel, and Moss (26) cited earlier in connection with the finding that no significant differences obtain between solutions written for the "Mined-Road" problem at four different levels of remoteness from reality. It was found also that the written solutions by groups were significantly superior to those written by individuals. Lorge *et al.* suggested that the apparent superiority of team solutions over individual solutions may be because five team members can and do ask more questions than do single individuals. Having collected more information, they are also able to advance more fruitful hypotheses about possible solutions.

In a later (53) study, Tuckman and Lorge formed 70 individuals who had already solved the "Mined-Road" problem into 14 *ad hoc* groups. The same problem was then posed to *S*s for group solution, for the purpose of confirming the hypothesized reasons for team superiority which emerged from the study (26) described above. The results showed that groups were superior to individuals, not necessarily because of the greater effectiveness of the groups for problem solving, but more likely because of the increased probability of getting a good solution from at least one member of a group of five men than from an individual.

Triandis, Bass, Ewen, and Mikesell (52) were interested in determining how much influence the creativity of individual members had on the creativity of the group as a whole. Using dyads, they had *S*s write solutions to problems (e.g., the "Frame and Immortality" problem) as individuals and as members of a dyad. The results showed that, at least for the dyads studied, as much as 72 per cent of the variance in group creativity could be predicted from knowledge of the individual creative abilities of the two members. Multiple correlations utilizing the interaction between ability scores of the two members did not improve the prediction of group performance. In one of the three experiments comprising the study, the relative "dominance" of the two dyad members was determined, and it was found that the performance of the dominant *S*s correlated higher with group performance than with the performance of the nondominant *S*s.

An interesting result of the study designed to investigate the adequacy of written reports of individuals and groups in problem-solving (27) was that, in contrast to records by trained observers, teams wrote solutions that were inferior to their actual performance on the "Mined-Road" problem constructed in reality, while individuals wrote solutions somewhat superior to their actual performance. Lorge and his associates suggested that, from a practical viewpoint, teams may need training to learn how best to present their pooled ideas and actions. It was suggested that the type of underreporting in the written solutions of groups, found in this study, may lead to overestimation of the actual functional behavior of individuals in studies comparing group contributions with individual contributions.

These studies do not seem to have considered the possibility that the major contribution of groups as opposed to individuals may be contained in the "originality" criterion for creativity. One of the conclusions of the Shaw study (44) was that groups were superior to individuals, owing to their greater rejections of unsuitable solutions. Applying this idea to the area of problem solving, we might say that groups discard more solutions which are not original than do individuals, since what may indeed be an original solution to one person in the group may not be for others—i.e., they may know of instances in which it has been tried before. This same line of reasoning might be applied to another criterion of creativity, that of "social usefulness." In brainstorming, this determination would, of course, come in the idea evaluation phase, after the idea production phase has been completed.

E. Leadership

The leader of a creative group is obviously just as important as the leader in any other group situation. In other fields of problem solving, leadership studies are not difficult to find, but in the particular area of group creativity they are rather rare.

Anderson and Fiedler (1) compared the creativity of four-man groups under two conditions of leadership. In half of the groups the leader directed the group discussions and contributed to the solution. In the other groups, leaders directed and guided group discussions and could encourage the rejection of ideas, but were not allowed to contribute their own ideas to the solution. The results, as noted above, were largely task specific, but there were some general indications that groups with participatory leaders were superior in quantity of output, while the groups with nonparticipatory leaders were superior in the quality of their product. It was suggested that groups with participatory leadership probably produced a greater quantity of solutions because they had one more member to contribute ideas. On the other hand, the groups with supervisory leaders probably produced higher quality solutions because of the screening and censoring function of the supervisory leader's role.

In a study which tested the effects of deferred judgment instructions against the textbook order of problem-solving discussion (problem-criterion-solution-evaluation), Brilhart and Jochem (2) found the deferred judgment pattern of problem-solving discussion to be significantly superior to the textbook pattern of problem solving. On the basis of these results, they suggested that there may be an advantage to be gained by training discussion leaders to separate ideation and evaluation.

Maier and Solem (30) were interested in the question of how a discussion leader might contribute to the quality of group thinking, if in fact, he does. To investigate this question they supplied 34 groups with leaders who were instructed to conduct a discussion, while another 33 groups were supplied with persons who were instructed simply to listen while the others discussed the problem. Both leaders and observers were asked to refrain from contributing their own views, and both sets of groups received the same problem. Both sets of groups gave similar answers to the problem before discussion but, after discussion, significantly more "leader" groups than "observer" groups switched to the correct answer. The authors suggested that the reason for this "leader" group superiority may be that individuals with minority opinions exerted more influence in the "leader" groups than in the "observer" groups. In the "leader" groups, when one person had the answer he was better able to convince the others than was true in "observer" groups. The problem that they used did not require a creative solution, but there appears to be some important implications here for groups trying to find a creative solution to a problem.

A study by Hoffman, Harburg, and Maier (20) tested the hypothesis that conflict, generated by group members who oppose the opinion advanced by a person in the group with authority (in this case, the foreman in the "Change of Work Procedures" problem) can increase the probability of high quality solutions. Subjects were divided into groups that included a foreman who was either considerate or dominant, and subordinate workers who were either weakly or strongly committed to the *status quo*. The results indicated that more creative solutions emerged when an authoritative leader (foreman) met with strong opposition from among the group members (workers). Hoffman *et al.* suggested that the differences in opinion promoted the creative problem solving by turning a choice situation into a problem situation.

The attitude which the leader has toward the problem to be solved was demonstrated by Maier and Solem (31) to be a crucial variable. The only difference between their experimental groups and control groups was that the leaders in the experimental groups were given instructions that were intended to make the leader perceive the task ("Change of Work Procedures" problem) as a problem to be solved, rather than as a decision-making situation. It was found that the experimental groups produced more than three times as many integrative solutions as the control groups, whose leaders did not have the experimental instructions. The integrative solutions were creative ways of satisfying what at first appeared to be opposed interests. Maier and Solem interpreted their results to mean that the number of alternative solutions and the quality of group performance can be increased by procedures that retard speedy decisions and cause the group to look at the situation as a problem to be solved. The above authors saw a need to create new solutions, in contrast to merely making choices from among already existing solutions.

Fiedler, Bass, and Fiedler (9), using a scale which indicates a leader's esteem for his least-preferred co-worker (LPC) as a measure of interpersonal attitude, investigated the relationships among leader attitude, group climate, and group creativity. Perceived group climate was assessed by means of some

"Group Atmosphere" scales. Fndings from this and a series of previous studies indicated that a particular type of leadership was not equally effective in all situations. Results suggested that groups under stress require a firm leadership technique, while groups with a relaxed, nonstressful climate function most successfully with a permissive and emotionally expressive leader. Subsequent studies (7, 37) have provided additional support for these findings. There has been a consistent relationship in all of these studies, between leader attitudes toward their least-preferred co-worker, group performance, and amount of stress shown by the group.

F. Training and Group Creativity

Cohen, Whitmyre, and Funk (5) studied the effects of training and cohesiveness on group creative thinking. They formed three basic groups: cohesive, noncohesive, and "nominal." Half of each of these groups was trained in a 10-hour course in creative thinking. The cohesiveness of the groups was determined by sociometric choices. The results showed the trained cohesive groups to be better than all other groups in terms of number of unique ideas produced for an ego-involving problem (the "Discharge" problem). There were no significant differences between the trained and the untrained noncohesive groups.

An experiment was designed by Meadow and Parnes (34) to evaluate the effectiveness of a course in creative problem solving. They administered a battery of tests to matched experimental and control groups at the beginning and the end of the course. The results showed the experimental group (with the training) to be superior to the control group in two measures of quantity of ideas and three out of five measures of quality of ideas. Meadow and Parnes interpreted these results as providing support for the training course as a means of producing increments in ability measures associated with practical creativity.

In a later study, Parnes and Meadow (43) evaluated the persistence of the effects of a course in creative problem solving. This time there were three groups: a group which had just completed the course, a group which had completed the course eight months prior, and a control group which had no training. The groups were matched for intelligence on the basis of WAIS scores. The problems used included the Guilford "Unusual Uses" and "Plot Titles" tests and the "Hanger" and "Broom" problems. The results showed the two trained groups to be so similar that they were pooled together for analysis. On all tests the trained groups performed significantly better than the control group. Thus it was concluded that training does cause a difference in creative ability and that this learned ability does last.

Further evaluation of the effectiveness of training was accomplished in another study by Parnes and Meadow (42). They studied the effects of "brainstorming" instructions on the creative problem solving of both trained and untrained *S*s. The results showed the groups which had been trained in a creative problem-solving course to be significantly better than the untrained groups in terms of number of good quality ideas. They also found that more good quality ideas were produced with brainstorming instructions than with nonbrainstorming instructions, but this apparently was for the trained and untrained groups combined. It was not reported whether or not there was a differential effect of the brainstorming instructions for the trained *vs.* untrained groups.

A study by Weisskopf-Joelson and Eliseo (54) showed their groups with "brainstorming" instructions to produce significantly more ideas than did groups with nonbrainstorming instructions, while the quality of ideas remained as good or better. However, neither of their groups had received any training in creative problem solving. These experimenters simply started the problem-solving sessions for the noncritical groups with brainstorming instructions.

In the two studies which found brainstorming in individuals to be superior to brainstorming in groups (6, 49), none of the *S*s received more than one hour of training in the brainstorming technique. Brilhart and Jochem (2) suggested on the basis of their study, discussed above, that training discussion leaders in the brainstorming device of separating ideation and evaluation would be advantageous.

A challenge to the concept of training creativity was offered by Parloff and Handlon (40), who suggested the interesting possibility that training in "creativity" may simply alter the standards of *S* so that they conform to those of the teacher, rather than bringing out anything really new in creative ability.

Chorness (4) described a study in which an attempt was made to determine whether the brainstorming technique would generate a significant increase in the number of creative responses within groups. The experimental groups were taught a course in leadership, making use of the brainstorming technique. The control groups were instructed "in the normal way." Both groups had been pretested on the Guilford creativity tests, and were tested again at the end of the experimental courses. No significant differences were found between the groups but, as Chorness pointed out, there is a vast difference between the responses made under group brainstorming conditions and the responses made on a pencil-and-paper test, in isolation, and under timed conditions.

Another study was described by Chorness (4) which took place in conjunction with the development of a creativity workshop. Utilizing exercises calculated to produce a strong awareness of the operational importance of the Guilford dimensions of creativity—i.e., flexibility, originality, problem sensitivity, etc., in real problems, another series of training sessions was arranged. The brainstorming technique was utilized "as a stimulant when required." This time the experimental groups performed significantly better than the control groups in terms of original responses, and in terms of a high ratio of original to total responses. It was concluded that the creativity workshop may be a good way to increase creativity in groups.

The divergent views, reported above, on the effectiveness of training for increasing the creativity of problem-solving groups should discourage the drawing of conclusions one way or the other. Certainly, enough concrete evidence has accumulated to encourage further attempts at the use of training, but more studies which demonstrate that a clear advantage is to be gained from training are obviously needed.

G. Group Characteristics

1. *Group Size*

Gibb (12) was interested in the effects of group size on creative problem solving, and studied various group sizes ranging from two to 96. He found that the number of ideas produced by the group increased as a negatively accelerating function of the size of the group. Threat-reducing instructions significantly increased the number and quality of solutions.

Taylor and Faust (48) studied problem-solving efficiency (on the "Twenty Questions" type of problem) in relation to group size. They used 15 individuals, 15 groups of two, and 15 groups of four. The results showed the group performance to be superior to individual performance on three of their efficiency measures, but the performance of groups of four was not superior to that of groups of two, except in number of failures to reach solution. On one score, man-minutes to reach solution, individuals were superior to groups, and groups of two were superior to groups of four.

Thomas and Fink (50), in a review article on the effects of group size, concluded that there have been too many methodological weaknesses in the studies which have been directed at determining the effects of group size, but that size must nonetheless be taken into consideration in any theory of group behavior.

In an anecdotal article, Isenson (21) reported that his organization (Army Research Office), as a matter of policy, successfully employs group creativity potentials to deal with many problems which face the army. Never more than six or eight members are employed in groups because it has been found from experience that a greater number of members tends to detract considerably from the desired close interplay.

2. *Intermember Relationships*

Hoffman (16) was concerned with the effect which homogeneity of member personalities might have on group problem solving. Using the Guilford-Zimmerman Temperament Survey (GZTS), he established groups with high positive profile correlations (homogeneous) and groups with zero or negative profile correlations (nonhomogeneous). In solutions to the "Mined-Road" (M-R) problem, results showed the nonhomogeneous groups to be superior to the homogeneous groups, and in solutions to the "Change of Work Procedures" (CWP) problem there was a tendency for the nonhomogeneous groups to produce more inventive solutions. Hoffman considered the CWP problem to involve both quality and acceptance, while the M-R problem involved only quality.

In order to test the generality of Hoffman's (16) earlier results, Hoffman and Maier (18) employed a widened variety of problems. Hoffman and Maier again utilized the GZTS to form the homogeneous and nonhomogeneous groups, but this time used the "Student Assistance Fund," "Painter-Inspector Argument," and "Point Distribution" problems as well as the previously used "Mined-Road" and "Change of Work Procedures" problems. Four of the problems had some component of quality and one ("Point Distribution") involved only acceptance. The results showed that the heterogeneous groups produced a higher proportion of high quality solutions than did the homogeneous groups on three of the four problems with quality components. Mixed-sex groups showed a tendency to produce higher quality solutions than did the all-male groups.

In another study utilizing the GZTS to form homogeneous and nonhomogeneous groups, Hoffman and Smith (19) found that homogeneous groups engaged more in promoting personal relations than did the nonhomogeneous groups, but that there were no other significantly different behaviors for the two groups. A study by Mellinger (36) may be related to these findings. He reported that the way people felt about each other affected their communication pattern. His experimental two-man groups were divided on the basis of "high mutual trust" and "low mutual trust." The "low mutual trust" groups spent more time in adjustive behavior, trying to reduce threat, and less time in creative-type behavior. His study, however, was not concerned with group creativity as such.

In one of a series of experiments (51), it was predicted that heterogeneous dyads whose members were given an opportunity to learn about each other would be more creative than would heterogeneous dyads that had no such opportunity. It was assumed that this knowledge about the other team member would reduce negative interpersonal relations. After 50 *S*s had judged 20 heterogeneous concepts (e.g., "socialized medicine," "war," etc.) on an 18-scale semantic differential, groups with differing degrees of homogeneity were formed on the basis of these semantic differential ratings—e.g., high cognitive-similarity groups were formed by pairing two extreme liberals, two extreme conservatives, two moderate liberals, or two moderate conservatives. Half of the dyads in each cognitive-similarity condition were given the opportunity to learn about the other member of their dyad by letting each member examine the semantic differentials produced by the other member.

There were no significant effects on the creativity of the dyads produced by heterogeneous or homogeneous grouping itself, or even by letting members in homogeneous groups learn something about the other member of their dyad. In support of the hypothesis, however, the heterogeneous dyads with knowledge about the other group member were very creative.

In another study by Triandis *et al.* (51), factor-analytic scores from a variety of measures were utilized to form dyads that were either heterogeneous or homogeneous in creative abilities and attitudes. It was hypothesized that, assuming homogeneous abilities, heterogeneity of attitudes would be beneficial to the creativity of the dyad for the "Fame and Immortality" ("Fame") and "Church" problems. The results supported this hypothesis for the "Fame"—"Church" problem sequence, but not for the reverse problem sequence. After it was noted that the "Fame" problem would not tend to create antagonism due to differences in attitude, it was suggested that working first on the "Fame" problem may have developed an interpersonal attraction between the members of the dyad, and that this attraction probably prevented the development of antagonism on the "Church" problem, enabling them to be more creative on that problem as well.

Subsequently a third experiment (51) was designed to test the hypothesis that problem sequence would affect dyadic creativity. Specifically, it was hypothesized that dyads working in the "Fame"—"Church" sequence would be more creative than those working in the "Church"—"Fame" sequence. It was, secondly, hypothesized that dyads which were heterogeneous in attitudes would be more creative than would attitudinally homogeneous dyads working in the "Fame"—"Church" sequence, but not for the reverse sequence. Thirdly, it was hypothesized that the dyads working in the "Fame"—"Church" sequence would have higher interpersonal attraction than in the reverse sequence.

Performance on each problem was analyzed as a function of problem sequence. The results supported the first hypothesis for the "Fame" problem but not for the "Church" problem. The second hypothesis, that dyads which were heterogeneous in attitudes would be more creative than those which were homogeneous, was also supported for the "Fame" problem but not for the "Church" problem. The third hypothesis (interpersonal attraction) was not supported. Triandis and his associates suggested that, since interpersonal attraction could not account for the results produced by the different problem sequences, it was possible that the "Fame" problem, because large numbers of easily produced responses were available, produced a "set" which was conducive to a high rate of creativity. Once adopted, this set then carried over to the "Church" problem. On the other hand, those dyads which worked first on the "Church" problem, for which solutions were not so easily created, may have acquired an inhibiting set.

Calvin, Hoffman, and Harden (3), using the "Twenty Questions" type of problem, studied the relationship between social atmosphere and intelligence in group problem solving. On the basis of ACE scores, they divided *S*s into "Bright" and "Dull" groups. It was hypothesized that "Bright" *S*s would function best in a permissive atmosphere, while "Dull" *S*s would function best in an authoritarian atmosphere. In the first of a series of three separate experiments, these experimenters found that "Authoritarian-Dull" groups performed significantly better than "Permissive-Dull" groups, but that there was no significant difference between the "Permissive-Bright" and the "Authoritarian-Bright" groups. The other two experiments, essentially replications with slight changes, found no statistically significant differences, but the trends supported the results of the first experiment.

As reported earlier, Fiedler, Bass, and Fiedler (9) found group creativity to be the resultant of complexly interacting group processes; also, that groups under stressful conditions required firm leadership, while groups working under pleasant relaxed conditions functioned better with permissive nondirective leadership.

Fiedler, Meuwese, and Oonk (10) conducted an investigation in the Netherlands, utilizing Catholic and Calvinist students. Each student participated as a member both of a homogeneous group of four members and a heterogeneous group consisting of two men from each religion. The experimenters appointed the chairman in half of the groups, while the other groups worked as "informal" teams. A measure of intelligence was obtained from a 14-item "analogies" test, and all of the *S*s completed some scales designed to measure interpersonal perception—prior to working in groups. For analysis of the data, the results of sociometric questions were used to divide the groups into "pleasant" relaxed cohesive groups and "unpleasant" tense uncohesive groups. The leaders of the "informal" groups were also inferred from sociometric questions—e.g., "Which of the group members had most influence on the opinions of others?" The results, which indicated no significant differences due to religious heterogeneity or homogeneity, showed that the intelligence of group leaders produced the greatest influence on group performance in pleasant cohesive groups. Fiedler and Meuwese (8) analyzed this study along with three other studies and arrived at a generalization that cohesive groups probably freed the leader from the effort needed to maintain the uncohesive type of group, and left him in a position to influence more effectively the performance level of the group by direct contribution to the solution of the problem.

Hoffman *et al.* (20) found conflict, resulting from the resistance of subordinate (worker) members to a solution by a member in an authority position (foreman), to increase the frequency of high quality solutions. It had also been hypothesized that more high quality solutions would be produced in mixed-sex groups than in one-sex groups, but this hypothesis was supported only suggestively.

Maier and Hoffman (29) have suggested that organizations which are run by a small group at the top of the organizational structure may be neglecting to use considerable creative potential which is available to them from lower levels in the organization. In order to vary the degree of interest in business among groups, they developed four different sets of groups. One set of groups was made up of actual employees of an industrial organization. The other three sets consisted of (*a*) upper division and graduate students in business and public administration (BPA) (business committed), (*b*) upper division students enrolled in a "psychology of human relations" course (slightly interested in business), and (*c*) lower division students in an introductory psychology course (not particularly interested in business at all).

All four groups worked on the "Change of Work Procedures" problem. The results showed a consistent decrease in the number of creative solutions as the degree of identification with the business world increased. This was interpreted to mean that problem-solving orientation produced more creative solutions than did authority-relation orientation. Maier and Hoffman contended that, since the BPA groups produced significantly fewer creative solutions than did the psychology classes, industry may be attracting people who will fit nicely into authoritarian setups, rather than people who will be creative.

The effect of feedback on problem-solving efficiency was investigated by Smith and Kight (45). Three types of groups were formed with 11-13 members in each group. In condition "A" (the experimental groups), small feedback groups of three or four members were set up that were instructed to describe and discuss the problem-solving behavior of each member. Condition "B" groups were broken up in the same way, but the subgroups discussed extraneous material instead of each member's problem-solving behavior. Condition "C" groups had no subgroups and received no feedback. The results showed that feedback did significantly increase the problem-solving efficiency of the experimental groups (at least on the "Twenty Questions" type of problem).

Yamamoto (55) investigated the relationship between sociometric status and creativity among adolescents. Students from Grades 7 through 12 were classified into three groups on the basis of tests of creative thinking. The High Creativity group consisted of those whose total creativity score was in the upper one-quarter of all scores. The Low Creativity group were those in the lower one-quarter, and the Medium Creativity group were those in the middle half. Peer nominations on the basis of six criteria—fluency, originality, flexibility, inventiveness, elaboration, and friendship—were then compared with the three test groupings. Among the junior-high-school students, the groupings from the two different sources tended to coincide, but among senior-high-school students only the boys tended to nominate peers in the same direction as had been determined by the creativity tests. Yamamoto suggested that this difference may have been due to a transition from a "psychological" criterion for the evaluation of creativity to a "social" criterion. He also suggested that a sex-typing process may have been involved.

Ziller, Behringer, and Goodchilds (56) varied group stability and group history of success and failure, in order to study their effects on group creativity. There were four variations of group stability: the addition, removal, or replacement of one group member; and a control group in which the original group remained the same. Initially the experimental groups were composed of two, three, or four members, but when the stability conditions were effected all groups contained three members. To induce histories of success or failure, prior to the stability condition changes, the groups were presented a numerosity task and a Meier Art Judgment item, after which the shifts in group size took place. Those groups which were to have "failure" histories were told that they had performed very poorly on those two tasks, while the "success" groups were told that they had done very well.

After group success or failure had been induced and the four stability conditions had been effected, the groups were instructed to compose as many captions as possible for a Saturday Evening Post cartoon. The results showed the open groups (those which experienced membership change) to be more creative than the closed groups. While the success-failure histories did not produce significant differences in the creativity of the groups, it was found that groups with a "failure" history evaluated their performance significantly lower than did "successful" groups. In comparison with "successful" groups, the regular members of "failure" groups also reported that the sessions were less "stimulating" and less "fun."

By way of summarizing the studies concerned with group characteristics, we might try to describe the characteristics of an ideal group structure for creative problem solving. It should be nonhomogeneous in terms of member personalities (16, 18) and in terms of attitudes (51), cohesive in terms of sociometric choices (9), and possibly should also be capable of producing some conflict between subordinate members and a member in an authority position (20). The group should be open (56) and, if the group is working under stressful conditions, the leader should be firm; but if the group is working under pleasant and relaxed conditions, the leader should be permissive and nondirective (9).

H. SUMMARY

After a consideration of the meaning of "creativity," the authors have briefly described the problems used for experiments in the group creative problem-solving process. The studies reviewed were related in terms of five basic variables that have been studied: (*a*) task variables, (*b*) individual vs. group, (*c*) leadership, (*d*) training, and (*e*) group characteristics. The latter category was subdivided into 1) group size and 2) intermember relationships.

Task variables were found to be an extremely important factor in experiments on group creative problem solving. In experiments that compared individuals to groups in creative problem-solving ability, the evidence tipped the scales in favor of the groups, though not on man-hour measures.

Those studies that have been concerned with the leadership form of creative problem-solving groups have all indicated that it is a significant variable. One form of leadership does not function optimally for all groups and all conditions. Studies that have evaluated the effectiveness of training for increasing group creativity have yielded varying results, with the majority of evidence in favor of training. Very little evidence is available on the relationship of group size to creativity. Characteristics of creative groups (heterogeneity, cohesiveness, openness, etc.) were reviewed and summarized.

REFERENCES

1. Anderson, L. R., & Fiedler, F. E. The effect of participatory and supervisory leadership on group creativity. *J. Appl. Psychol.*, 1964, **48**, 227-236.
2. Bellhart, J. K., & Jochem, L. M. Effects of different patterns on outcomes of problem-solving discussion. *J. Appl. Psychol.*, 1964, **48** (3), 175-179.
3. Calvin, A. D., Hoffman, F. K., & Harden, E. L. The effect of intelligence and social atmosphere on group problem solving behavior. *J. Soc. Psychol.*, 1957, **45**, 61-74.
4. Chorness, M. H. Increasing creativity in problem-solving groups. *J. Communication*, 1958, **8**, 16-23.
5. Cohen, D., Whitmyre, J. W., & Funk, W. H. Effect of group cohesiveness and training upon creative thinking. *J. Appl. Psychol.*, 1960, **44**, 319-322.
6. Dunnette, M. D., Campbell, J., & Jaastad, K. The effect of group participation on brainstorming effectiveness for two industrial samples. *J. Appl. Psychol.*, 1963, **47**, 30-37.
7. Fiedler, F. E. Leader attitudes, group climate, and group creativity. *J. Abn. & Soc. Psychol.*, 1962, **65** (5), 308-318.
8. Fiedler, F. E., & Meuwese, W. A. T. Leader's contribution to task performance in cohesive and uncohesive groups. *J. Abn. & Soc. Psychol.*, 1963, **67** (1), 33-47.
9. Fiedler, F. E., Bass, A. R., & Fiedler, J. M. The leader's perception of co-workers, group climate and group creativity: A cross-validation. Technical Report No. 1, Group Effectiveness Laboratory, University of Illinois, Urbana, Illinois, May, 1961.
10. Fiedler, F. E., Meuwese, W. A. T., & Oonk, S. An exploratory study of group creativity in laboratory tasks. *Acta Psychol.* (*Amsterdam*), 1961, **18**, 100-119.
11. Flanagan, J. C. The definition and measurement of ingenuity. In C. W. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development.* New York: Wiley, 1963.
12. Gibb, J. R. Effects of group size and threat reduction upon creativity in a problem-solving situation. *Amer. Psychologist*, 1951, **6**, 324 (Abstract).
13. Golann, S. E. Psychological study of creativity. *Psychol. Bull.*, 1963, **60**, 548-565.
14. Haefele, J. W. Creativity and Innovation. New York: Reinhold, 1962.
15. Hitt, W. D. Toward a two-factor theory of creativity. *Psychol. Rec.*, 1965, **15** (1), 127-132.
16. Hoffman, L. R. Homogeneity of member personality and its effect on group problem solving. *J. Abn. & Soc. Psychol.*, 1959, **58**, 27-32.
17. ———. Conditions for creative problem solving. *J. of Psychol.*, 1961, **52**, 429-444.
18. Hoffman, L. R., & Maier, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *J. Abn. & Soc. Psychol.*, 1961, **62**, 401-407.
19. Hoffman, L. R., & Smith, G. G. Some factors affecting the behaviors of members of problem-solving groups. *Sociometry*, 1960, **23**, 273-291.

20. Hoffman, L. R., Harburg, E., & Maier, N. R. F. Differences and disagreement as factors in creative problem solving. *J. Abn. & Soc. Psychol.*, 1962, **64**, 206-214.
21. Isenson, R. S. Group creativity: A research planning tool. *Person. Admin.*, 1964, **27** (6), 36-38.
22. Krech, D. E., & Crutchfield, R. S. Elements of Psychology. New York: Knopf, 1958.
23. Lorge, I., & Solomon, H. Two models of group behavior in the solution of Eureka-type problems. *Psychometrika,* 1955, **20**, 139-149.
24. ———. Individual performance and group performance in problem solving related to group size and previous exposure to the problem. *J. of Psychol.*, 1959, **48**, 107-114.
25. ———. Group and individual performance in problem solving related to previous exposure to problem, level of aspiration, and group size. *Behav. Sci.*, 1960, **5**, 28-38.
26. Lorge, I., Tuckman, J., Aikman, L., Spiegel, J., & Moss, G. Solutions by teams and by individuals to a field problem at different levels of reality. *J. Educ. Psychol.*, 1955, **46**, 17-24.
27. ———. The adequacy of written reports in problem solving by teams and by individuals. *J. Soc. Psychol.*, 1956, **43**, 65-74.
28. Maier, N. R. F., & Hoffman, L. R. Quality of first and second solutions in group problem solving. *J. Appl. Psychol.*, 1960, **44**, 278-283.
29. ———. Organization and creative problem solving. *J. Appl. Psychol.*, 1961, **45**, 277-280.
30. Maier, N. R. F., & Solem, A. R. The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions. *Hum. Relat.*, 1952, **5** (3), 277-288.
31. ———. Improving solutions by turning choice situations into problems. *Person. Psychol.*, 1962, **15**, 151-157.
32. Marquart, D. I. Group problem solving. *J. Soc. Psychol.*, 1955, **41**, 103-113.
33. McPherson, J. H. A proposal for establishing ultimate criteria for measuring creative output. In C. Taylor & F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development.* New York: Wiley, 1963.
34. Meadow, A., & Parnes, S. J. Evaluation of training in creative problem solving. *J. Appl. Psychol.*, 1959, **43** (3), 189-194.
35. Meadow, A., Parnes, S. J., & Reese, H. Influence of brainstorming instructions and problem sequence on a creative problem-solving test. *J. Appl. Psychol.*, 1959, **43**, 413-416.
36. Mellinger, G. D. Interpersonal trust as a factor in communication. *J. Abn. & Soc. Psychol.*, 1956, **52**, 304-309.
37. Meuwese, W. A. T., & Fiedler, F. E. Leadership and group creativity under varying conditions of stress. Technical Report No. 22, Group Effectiveness Laboratory, University of Illinois, Urbana, Illinois, March, 1965.
38. Moos, R. H., & Speisman, J. C. Group compatibility and productivity. *J. Abn. & Soc. Psychol.*, 1962, **65**, 190-196.
39. Osborn, A. F. Applied Imagination. New York: Scribner's, 1963.
40. Parloff, M. B., & Handlon, J. H. The influence of criticalness on problem solving in dyads. *Psychiatry,* 1964, **27** (1), 17-27.
41. Parnes, S. J. Effects of extended effort in creative problem solving. *J. Educ. Psychol.*, 1961, **52**, 117-122.
42. Parnes, S. J., & Meadow, A. Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *J. Educ. Psychol.*, 1959, **50**, 171-176.

43. ———. Evaluation of persistence of effects produced by a creative problem-solving course. *Psychol. Rep.*, 1960, **7**, 357-361.

44. SHAW, M. E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, **44**, 491-504.

45. SMITH, E. E., & KIGHT, S. S. Effects of feedback on insight and problem-solving efficiency in training groups. *J. Appl. Psychol.*, 1959, **43**, 209-211.

46. SPRECHER, T. B. A proposal for identifying the meaning of creativity. In C. W. Taylor and F. Barron (Eds.), *Scientific Creativity: Its Recognition and Development.* New York: Wiley, 1963.

47. TAYLOR, C. W., & BARRON, F., *Eds.* Scientific Creativity: Its Recognition and Development. New York: Wiley, 1963.

48. TAYLOR, D. W., & FAUST, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. *J. Exper. Psychol.*, 1952, **44**, 360-368.

49. TAYLOR, D. W., BERRY, P. C., & BLOCK, C. H. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Admin. Sci. Quart.*, 1958, **3**, 23-47.

50. THOMAS, E. J., & FINK, C. F. Effects of group size. *Psychol. Bull.*, 1963, **60**, 371-384.

51. TRIANDIS, H. C., HALL, E. R., & EWEN, R. B. Member heterogeneity and dyadic creativity. Office of Naval Research contract report NR 177-472, Nonr-1834 (36), University of Illinois, Urbana, Illinois, 1964.

52. TRIANDIS, H. C., BASS, A. R., EWEN, R. B., & MIKESELL, E. H. Team creativity as a function of the creativity of the members. *J. Appl. Psychol.*, 1963, **47** (2), 104-110.

53. TUCKMAN, J., & LORGE, I. Individual ability as a determinant of group superiority. *Hum. Relat.*, 1962, **15**, 45-51.

54. WEISSKOPF-JOELSON, E., & ELISEO, T. S. An experimental study of the effectiveness of brainstorming. *J. Appl. Psychol.*, 1961, **45**, 45-49.

55. YAMAMOTO, K. Creativity and sociometric choice among adolescents. *J. Soc. Psychol.*, 1964, **64**, 249-261.

56. ZILLER, R. C., BEHRINGER, R. D., & GOODCHILDS, J. D. Group creativity under conditions of success or failure and variations in group stability. *J. Appl. Psychol.*, 1962, **46**, 43-49.

Department of Psychology
The University of Arizona
Tucson, Arizona 85721

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

March, 1966
Volume 62, Second Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

MARCH, 1966

(Manuscripts are printed in the order of final acceptance)

The problem of temporal order in stimulation and perception 141
By James J. Gibson

Electrocutaneous pain thresholds in humans to low frequency square-wave pulses 151
By Robert Plutchik and Henry Bender

Structure of maternal attitudes in two cultures 155
By George M. Guthrie

The anatomy of abnormality 167
By William G. Herron and Robert E. Kantor

A note on Smart's study of birth rank and affiliation in male university students 177
By Jan E. E. de Lint

Psychological needs and verbally expressed aggression of adolescent delinquent boys 179
By John E. Horrocks and Nathan W. Gottfried

(OVER)

Copyright © 1966 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved
Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

Potential value of educational background data in the selection and classification of military personnel 195
By Chester J. Judy

A note about communication research 201
By Daniel K. Stewart

Raven's Progressive Matrices: Construct validity 205
By William C. Bingham, Henry R. Burke, and Stewart Murray

The effect of verbal suggestion on the reversal rate of the Ames trapezoid illusion 211
By Margaret Kathleen Cappone

The relation of verbal intelligence and color-form discriminative ability to children's color-matching and form-matching behavior 221
By Norman L. Corah, Sally Ann Jones, and Barbara B. Miller

Personality changes in psychiatric residents during training 229
By William S. Kogan, Erling E. Boe, Edward F. Gocka, and Merlin H. Johnson

Distribution of anagram solution times 241
By Ira T. Kaplan, Thomas Carvellas, and Barbara J. Breinin

A note on two basic forms of concepts and concept learning 249
By Gary A. Davis

A study of certain visual effects occasioned by factors of so-called glare . . 255
By Richard F. Haines and S. Howard Bartley

Loudness adaptation as a function of frequency, intensity, and time . . 267
By Warren H. Teichner and Ernest Sadler

Category clustering in free recall 279
By John A. Robinson

Cultural and psychological differences among occupational groups . . . 287
By E. K. Eric Gunderson and Jack L. Mahan

Published as a separate and in *The Journal of Psychology*, 1966, 62, 141-149.

THE PROBLEM OF TEMPORAL ORDER IN STIMULATION AND PERCEPTION*[1]

Graduate Psychological Laboratories, Cornell University

JAMES J. GIBSON

A. INTRODUCTION

Between 1950 and 1954 Lashley wrote three related papers entitled "The Problem of Serial Order in Behavior" (6), "Dynamic Processes in Perception" (7), and "In Search of the Engram" (5). He was deeply concerned with the puzzle of time and memory, with the way in which temporal order enters into behavior, perception, and learning. He had been searching all his life for the physiological basis of memory—for the engram—but he had to conclude that he had discovered nothing as to its real nature and that no one else had either. The idea that the traces left by experience might be sensorimotor connections in the brain he liked no better than the older notion that traces were little images impressed on the neural tissue. Both of these ideas were inconsistent with the results of his experiments. "I sometimes feel," he said, "that the necessary conclusion is that learning just is not possible. It is difficult to conceive of a mechanism which can satisfy the conditions set for it" (5, p. 477). "It seems certain that the theory of . . . paths from sense organs to association areas to the motor cortex is false. . . . It is not possible to demonstrate the localization of a memory trace anywhere within the nervous system . . . and the so-called association areas are not storehouses for specific memories" (5, p. 478).

At the very end of the paper, however, Lashley sounded another note. He suggested that "the learning process must consist of the attunement of the elements of a complex system in such a way that a particular combination or pattern of cells responds more readily than before the experience" (5, p. 479). No one knows how this might occur, and the suggestion is vague, but note that this idea is novel. If learning is a kind of *resonance* in the nervous system, a tuning of the system to certain inputs, then it is not any sort of storage of engrams or depositing of traces. Let us follow up this

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 19, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Psi Chi address delivered at meetings of the Eastern Psychological Association, April 1965. Research on the study of sequential perception was supported in part by the Office of Naval Research under Contract 401(14) with Cornell University.

idea. It suggests to me an astonishing hypothesis, that learning does not necessarily depend on memory as it has always been conceived. Perhaps we should re-examine the problem of temporal order as Lashley started to do. Perhaps the mechanism of learning that proved so elusive can be found in the solution of that problem.

B. THE MUDDLE OF MEMORY

There is no satisfactory definition of memory that I can find, except the commonsense one that appeals to introspection—the faculty of remembering. It cannot be defined as the rearousal of past experience, for surely not all memory is a matter of conscious experience. It cannot be defined as any lasting modification of neural function, for that would include growth. It cannot be defined as the general dependence of the present on the past, for that confuses memory with causation. Experimental psychologists usually do not even try to define memory except in terms of the experiments they perform. Memory is what the memory methods are concerned with. But the trouble with this is that a great many experiments requiring the reproduction of items presented, or the recognition of them, could be and often are called experiments on perception instead of memory. The act of drawing a visual pattern, or matching it to a sample, may be taken either as form-perception or as memory for form. The objective operations do not distinguish memory from perception. Only our subjective feeling about them separates the two kinds of activity. We have the feeling that perception is confined to the present, whereas memory refers to the past. But this distinction, be it noted, is wholly introspective. Moreover, as will be evident later, it cannot be made with any clarity.

There is another way, you might suppose, in which the activity of perception could be separated from the activity of memory—namely, by assuming that all sensitivity to simultaneous inputs constitutes perception and all sensitivity to successive inputs constitutes memory. The idea that form or space is perceived, whereas sequence or time is remembered, lurks vaguely at the back of our psychological thinking; but it is surely mistaken. As I will try to show, the pattern and sequence of stimulation are too closely interwoven, space and time are too interdependent, for this formula to be true.

C. THE POSSIBLE SOURCES OF CONFUSION

1. *The Before-After Relation*

One reason for the muddle of memory, I suspect, is our failure to consider the problem of temporal order *objectively*. We have habitually conceived time as consisting of past, present, and future, instead of a dimension defined by the relation of prior to subsequent. The stream of consciousness, as William James described it, exhibits the travelling moment of present time, the feeling of "now," and this divides the stream into a past extending in one direction and a future extending in the other. But the stream of stimulation has no such unique moment. Neither does the flow of external events. A man is aware of *now* but a clock is not. Consequently, a man is aware of his "past" but the clock is not. All the clock does is conform to the relation of before-after. Objective time, consisting of the sequential order of events, has no past and no memory.

2. *The Fallacy of Assigning Perception to the Present and Memory to the Past*

We have been assuming, without ever examining the assumption, that memory applies to the past, perception is confined to the present, and the kind of activity we call expectation or prediction applies to the future. The separation of these activities, and of the physiological basis sought for each, has only an introspective criterion to justify it. The classification is questionable and I wonder if we can afford to keep them separate. Perception, memory, and expectation smell of faculties.

In the first place, an exact division between the present and the past by means of introspection has never been possible; and we, therefore, do not know when perception leaves off and memory begins. The travelling moment of present time is not a razor's edge, as James observed, for that would cease to exist; there is a "sensible" present, although no one has been able to measure its duration. I would add that we cannot be sure when perception stops and *expectation* begins, either, for a recent experiment at Cornell with an expanding silhouette in the field of view (8) strongly suggests that the perception of approach and the expectation of collision are not separate.

Even more cogent is the evidence for the perception of motion. A change of visual pattern in time or a displacement of one part of the array relative to the rest of it is perceived, not remembered; and a visible motion may go on for a long time, or may even be endless, like that of a whirlpool or a waterfall. Perception is an activity, not an instantaneous event shrunken to the razor's edge. The same is true of the perception of rhythm or any auditory sequence. The lower animals react to an optical motion as if it were a simple stimulus, where they would not react at all to a motionless pattern. Kinesthetic stimuli, like acoustic, must have duration to exist. Every possible sensation, in fact, by Titchener's criteria, must have the attribute of protensity.

3. *The Fallacy of Assigning Pattern to Perception and Sequence to Memory*

The perception of time has been recognized in psychology, or by some psychologists, but they have never faced up to the contradiction between this kind of time and memory-time, subjective time. We have accepted space-perception as a valid problem, but have been uncomfortable about time-perception. We have attempted to keep separate the problem of detecting patterns (objects) and that of detecting sequences (events). And hence the equivalence of pattern and sequence, of space and time, has seemed to be a puzzle which had better be swept under the rug than confronted.

For example, a series of items can be presented to a subject either in adjacent order (with an exposure device) or in successive order (with a memory drum); and the same groups are apprehended in the spatial distribution as in the temporal distribution. Is the first a span of perception and the second a span of memory? The apprehension is equivalent in both cases.

The eyes of primates and men work by scanning—that is, by pointing the foveas at the parts of the scene in succession. The eyes of rabbits and horses do not, for they see nearly all the way around at once and have retinas with little foveation. Does this mean that a horse can perceive his environment, whereas a man can apprehend it only with the aid of memory? I once thought so (2, p. 158 *ff.*) on the theory that successive retinal images must be integrated by memory, but this now seems to me wrong. It is truer to suppose that a visual system can substitute sequential vision for panoramic vision, time for space. Looking around is equivalent to seeing around, with the added advantage of being able to look closely. It is no harder for a brain to integrate a temporal arrangement than a spatial arrangement.

D. THE EXPERIMENTAL STUDY OF SEQUENTIAL PERCEPTION

My suspicions about the orthodox theory of memory are based on 15 years of research having to do with the perception of moving things or processes and with the control of locomotion. A good many of these experiments were unpublishable because they did not work out. But those that failed and those that worked both point to the general conclusion that the flow of stimulation, its transformation, reveals the activity of perception better than the unchanging form, the picture, that we generally use as a stimulus. A form, in fact, is nothing but a continuous nontransformation just as nonmotion is a special case of motion. Hence, the perception of space is incomprehensible unless we tackle it as the problem of space-time. Even the cues for depth, the unambiguous ones, are kinetic and not static. I mean by these the parallax information that comes from moving about in the world.

Parenthetically, let me say that the experiments began to be intelligible only when optical motion was clearly distinguished from material motion, the proximal stimulus from the distal source of stimulation, and only after it was realized how remarkably unlike the motion of an object the motion of light is. A perspective transformation is unlike a rotation and a magnification is unlike an approach, but each *corresponds* nevertheless to its objective event. I will not try to summarize these experiments here, but some of their implications are interesting.

We found that continuous optical transformations can yield quite simple perceptions, but that they yield two kinds of perception at the same time, one of change and one of nonchange (4). The perspective transformation of a rectangle, for example, was always perceived as both something rotating and something rectangular. This suggests that the transformation as such is one kind of stimulus information, for motion, and that the invariants under transformation are another kind of stimulus information, for the constant properties of the object. The constancy of the phenomenal object in such experiments, far from being destroyed by the variation of momentary sensation, may be enhanced by it (3). This hypothesis of a persisting stimulus underlying a flux of changing stimuli, of invariant variables accompanying the variant variables, is a very powerful one. I am tempted to call it a fact, since it is a familiar idea in mathematics, and we should have thought of it long ago. There is almost always some permanence imbedded in any change. Note particularly that this continuing stimulus, if it is registered, explains the phenomenal constancy and identity of the object without any appeal whatever to a cumulative memory process.

The study of pattern vision in men and animals has been going on for a long time, and the theoretical problems are familiar ones. The study of what I will call (for the moment) change-of-pattern vision is only beginning, and its theoretical problems are new. We may have to relinquish the old presuppositions in studying the new field. If, as is likely, transformations over time were the usual stimuli in the course of the evolution of visual systems, and nontransformations over time were rare and exceptional, then the primacy of form or pattern comes into question. There is no "form" left in a continuous transformation. It has vanished, and all that remains is the invariants. Form discrimination as we test it by the presentation of two stationary pictures is irrelevant. There are individual pictures in a strip of motion picture film but not in the visual system of an insect. It is no wonder, then, that a lower animal may have good motion discrimination with poor stationary pattern vision. Hints of this primitive sensitivity are still obtainable in our own pictorially educated vision. The afterimage from watching the expansion or contraction of the light from a rotating spiral is contraction or expansion respectively, and the direction of this apparent motion does not depend on the comparison of two different patterns at two different times. So far from motion being the comparison of two static pictures, it is the other way round—a static picture is only the null case of motion.

E. New Theorems

Not only in vision but also in active touch and obviously in audition, the natural input to the perceptual organ is a changing array of stimulation—not a stationary one. The registering of the change is one task of the nervous system, and the detection of the underlying permanence is another. Perhaps, as Lashley suggested, the brain resonates to whatever is invariant under transformation and becomes increasingly attuned to it with recurrence over time. If so, perception and learning could be accounted for without any assumption of memory considered as an accumulation of traces. The brain would be a self-tuning resonator, not a storehouse.

A number of theorems are implied by this theory of apprehension, and I will try to make them explicit.

1. A succession can be perceived without having to convert all its elements into a simultaneous complex.

As Boring points out (1, p. 576), psychologists have generally accepted the proposition that a succession of items can be grasped only if the earlier ones are "held over" so as to be combined with the later ones in a single composite. Hence comes the theory of traces, requiring that every percept begins a trace, that they accumulate, and that every memory trace can theoretically reinstate its percept. The whole line of reasoning can be pushed into absurdity. The alternative is to deny the proposition on which it is based.

2. Stimulation normally consists of *successivities* as well as *adjacencies,* and either will excite a receptive system.

This says that there are two kinds of order in stimulation, adjacent and successive (2, p. 63), the latter being on the same footing as the former. A visual margin between dark and light is a stimulus of the first sort, and a visual transient between dark and light is one of the second sort. The important fact in either case is the direction of the intensity step. An upward margin is never confused with a downward margin, and an "on" is never confused with an "off." The retina in fact contains anatomical units that detect both kinds of transition, the spatial and the temporal. The sequential detectors do not have to make a comparison of intensity *now* with the memory of intensity *then* (an *off* receptor surely does not need a memory), and similarly the adjacent detectors do not have to compare intensity above the margin with intensity below it (or to the right and to the left of it). Note that stimulation is asserted to be relational by definition and that sensations of absolute brightness are left out of account as being irrelevant.

3. The detection of *different-from-before* is simply an alternative to the detection of *same-as-before,* and both are primitive.

This says that the perception of a change or transformation is the reciprocal of a perception of nonchange or nontransformation. Neither requires an act of comparison. If the principle is extended over time, recognition can be explained as the detection of same-as-before without the requirement that the new percept be compared to the trace of the old one and be judged to match. Similarly the perception of novelty, the opposite of recognition, can be explained as the detection of different-from-before without having to assume that the memory is searched for traces of the past and no match is found.

The problem of phenomenal identity in perception experiments using sequential stimulation (e.g., moving shadows) is not fundamentally different from the problem of recognition in memory experiments using sequential stimulation.

4. With increasing elaboration of the successive order of stimulation and the adjacent order of stimulation, limits are finally reached to the apprehension of both successive units and adjacent units.

This says that the temporal structure of stimulation and the spatial structure are both grouped into units, sometimes hierarchical and sometimes unrelated units, and that apprehension can span only a limited number of these units. In short, apprehension spans time—more exactly, events in time—but this does not have to be conceived as a span of memory, or as "primary," "immediate," or "short-term" memory. Apprehension also spans objects in space, and this does not have to be conceived as exclusively a span of perception. It is as if the brain could register only so many different successive events occurring in the same place and only so many different objects existing at the same time in different places.

As an example of what is meant by *successive units* in the above theorem, consider the following. The last paragraph may be said to consist of 571 letters, or 112 words, or 26 phrases, or four sentences, or one thought. Which unit do we choose for a description of the source of stimulation? (The same question arises if we consider the voice instead of the writing.) Clearly, all are legitimate. Each kind of unit is nested within a superordinate unit, and each kind is composed of subordinate units. There is no clear end in either direction, either of higher or lower order. And the same principle of forms within forms applies to the events of the natural world as to language. There is a *grammar* of events and a corresponding grammar of stimulation. Since there are normally no absolute units with which a psychologist can rest content, the choice of units is not a matter of psychological theory. It is only a matter of their being appropriate for the level of apprehension being considered.

5. In this theory, the function of perceiving is to maintain the contact of an organism with its environment.

Since, for an active organism in an eventful world, the stimulus energy is a continual flux, the main problem of perceiving is to detect the information that specifies the objects of the world—that is, to register the invariants under transformation. This is a solution to the puzzle of perceptual constancy. More than that, it resolves certain perplexities in the theory of memory, for it implies that what we call memory and recognition is often only a special case of invariant-detection. *Recollection* is then a sort of human luxury, an incidental ability to contemplate the past. Learning can occur without the intervention of remembering, just as thought can be imageless, and just as perception, in my view, can be sensationless.

F. CONCLUSION

The puzzle of the detection of both change and permanence in the world goes back nearly 2500 years. Heraclitus and his followers argued that all is flux and permanence is illusory. Parmenides argued that the world is immutable and change is illusory. His pupil, Zeno, even argued that motion is illusory. The controversy was a standoff until it was supposedly resolved by the doctrine that the atomic elements explain the permanencies while their separation and recombination explain the changes. But this resolution is not satisfactory to a psychologist, for it does not apply very well to the problem of what the *information* might be for apprehending both change and permanence.

The French have a neat saying, *Plus ça change, plus c'est la même chose.* It asserts, I think, the reality of permanence within change, of invariants underlying transformation. More than that, it suggests that the identity of a thing, its constancy, can emerge in perception only when it is observed under changing circumstances in various aspects. The static form of a thing, its image or picture, is not at all what is permanent about it. A form frozen in time is ambiguous information and is not a typical stimulus for our receptive surfaces. Only when the perspectives flow can we notice the distinctive features of the solid object. The invariants can be detected only if there is opportunity to distinguish them from the variations. This is surely not a matter of storing up all the cross-sections of past experience and of performing mental operations on the composite brought up to date. Let us not treat memory like a sacred cow, just because we are sometimes aware of remembering.

G. Summary

The concept of memory in its relation to perception is a muddle. If we accept the fact of sequential perception, rejecting the fiction of momentary pattern-perception, matters become more intelligible. If perception involves the apprehension of a changing world, not a frozen one, the problem is that of detecting invariants under transformation. The permanence can be isolated just because the perspectives change. The latter do not have to be stored up and put together in a composite. The mechanism of perceptual learning is one in which the nervous system resonates to the invariants of the stimulus flow, as Lashley suspected, not one of storage and retrieval of engrams. The recalling of the past, the capacity (in some persons) to summon memory images into consciousness, may well be a quite incidental accompaniment of learning, not its basis.

References

1. Boring, E. G. Sensation and Perception in the History of Experimental Psychology. New York: Appleton-Century-Crofts, 1942.
2. Gibson, J. J. The Perception of the Visual World. Boston: Houghton Mifflin, 1950.
3. ———. The useful dimensions of sensitivity. *Amer. Psychol.*, 1963, **18**, 1-15.
4. Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *J. Exper. Psychol.*, 1957, **54**, 129-138.
5. Lashley, K. S. In search of the engram. In *Physiological Mechanisms in Animal Behaviour* (*Symp. Soc. Exper. Biol., No. 4*). New York: Academic Press, 1950.
6. ———. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley, 1951.
7. ———. Dynamic processes in perception. In J. F. Delafresnaye (Ed.), *Brain Mechanisms and Consciousness*. Oxford: Blackwell, 1954.
8. Schiff, W. Perception of impending collision: A study of visually directed avoidant behavior. *Psychol. Monog.*, 1965, **79**, Whole No. 604.

Graduate Psychological Laboratories
Building 3
Cornell University
Cornell Industry Research Park
Ithaca, New York

Published as a separate and in *The Journal of Psychology,* 1966, **62**, 151-154.

ELECTROCUTANEOUS PAIN THRESHOLDS IN HUMANS TO LOW FREQUENCY SQUARE-WAVE PULSES*[1]

Department of Psychology, Hofstra University; and Matrix Corporation

ROBERT PLUTCHIK AND HENRY BENDER

A. INTRODUCTION

It is well known that low frequency sensory inputs will produce EEG and other physiological changes (6). In addition, several studies have shown that stimuli at about 3 cps produce appreciably different responses from those at higher or lower frequencies (7). One hypothesis that has been suggested to account for these observations is that low frequency stimulus inputs interact with already existing body rhythms to create inhibitory or facilitatory effects. If this possibility is assumed, it would seem likely that electrocutaneous pain thresholds might also show such a phenomenon. To test this would require a study of pain thresholds in the low frequency range (i.e., the range comparable to most EEG frequencies) covering 1 to 15 pulses per second.

Although electrical stimulation is widely used in research, most studies utilize pulse rates above this range. Almost all the studies of the Virginia Cutaneous Project (11) used frequencies over 60 pps (e.g., 4), although Gilmer (2) presents some conclusions concerning the relevance of the low frequencies for cutaneous electro-pulse communication. Gibson (1) has reported that the pain threshold at the finger gradually decreases from about 2 ma to a little over 1 ma as the repetition rate increases from 10 pps to 100 pps. Thresholds depend also on pulse duration, with very brief pulses (on the order of one msec) having only touch qualities even at high intensity. In the light of these considerations, the present study attempted to determine pain thresholds in the range of 1-15 pps.

B. METHOD

Tektronix waveform (Type 162) and pulse generators (Type 161) were used to control the output of an Applegate Model 230 constant current

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 31, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by Contract Nonr-2252 (01) with the Office of Naval Research. Reproduction for any purpose of the United States Government is permitted.

stimulator so that each pulse had a uniform pulse width of 50 msec. Pulse trains were five seconds long and rest intervals were 30 seconds long; these sequences were controlled by a Dual-Trol Industrial Timer.

Twenty college students were used as subjects, 10 males and 10 females, ranging in age from 18 to 22 years. Silver disc EEG electrodes, one cm in diameter, were taped to the ventral pads of the first and fourth fingers of the right hand of each subject. No electrode paste was used. The subjects were then presented with rates of 1, 3, 6, 10, and 15 pulses per second in counterbalanced order, three times each, using a five-second period of stimulation and a 30-second interstimulus period. For each pulse rate, the intensity of the current was increased in discrete 0.1 ma increments in ascending order, using the method of limits. At the end of each five-second stimulus period, the subject reported whether or not the sensation was "just noticeably unpleasant" or "just noticeably painful." The latter point was taken as the pain threshold. The data obtained from the males and the females were treated separately.

C. Results and Discussion

Figure 1 presents the relationship obtained between threshold and pulse rate and shows clearly that the males have a consistently higher pain threshold than the females. Thresholds are highest at 1 pps and gradually decrease. The slope of the curve suggests that the threshold will probably continue to decline at pulse rates greater than 15 pps.

The actual amount of energy the subject receives at a pulse rate of 15 pps is many times greater than at 1 pps. The fact that the pain threshold is lower at 15 pps than at 1 pps suggests some sort of integration of electrical energy in the skin so that the perceived intensity of the stimulation is greater at the higher pulse rates. This implies that pain threshold is not determined by the energy in a single pulse, but by some function of pulse height, duration, and repetition rate—a conclusion also reported by Gilmer (2).

It is interesting to note that no unusual or discontinuous effects were observed at a pulse rate of 3 pps, in contrast to some of the studies cited earlier.

The second major finding is the existence of a clear-cut sex difference in pain threshold at all pulse rates, with almost no overlap of individuals. Although this could conceivably reflect certain cultural expectations, there is some reason to believe that it might indicate a physiologic-anatomic sex difference in skin characteristics. Sternbach and Tursky (10) have reported that the mean slope of the power function for electric shock obtained from a

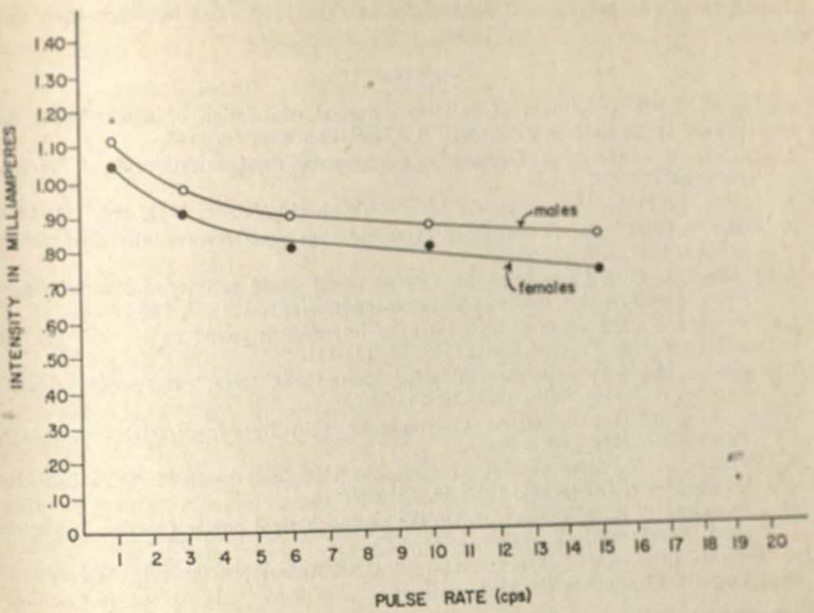

FIGURE 1
Mean Pain Thresholds for 10 Male and 10 Female Subjects as a Function of Pulse Rate

group of women is less than the slope obtained from a group of males. Green (3, p. 1304) has suggested that there seems to be a "direct relationship between the thickness of a subject's skin and his threshold." He implies that women should therefore have a lower threshold than should men.

In this context it is interesting that Shaklee (9) has reported that female rats have a lower pain threshold to electric shock than do male rats. This observation was made on all test days and with two different methods for measuring thresholds.

Also consistent with this interpretation is the fact that another skin property—skin resistance—shows a highly significant difference in mean level between males and females (8). This difference is probably related to some aspects of skin structure rather than sweating rates, since it has been shown that racial differences in skin resistance exist that are not highly correlated with amount of sweating (5). These investigators also noted that no racial differences were found in heart rate, respiratory rate, finger-skin temperature, or blood pressure. This further supports the possibility that sex differences in pain threshold reflect real differences in skin properties between men and women.

REFERENCES

1. GIBSON, R. H. Conditions of painless electrical stimulation of touch. Paper delivered at the Eastern Psychological Association meeting, 1962.
2. GILMER, B. v. H. Toward cutaneous electro-pulse communication. *J. of Psychol.*, 1961, **52**, 211-222.
3. GREEN, R. T. Absolute threshold for electric shock. *Nature,* 1962, **194**, 1303-1304.
4. HAHN, J. F. Cutaneous vibratory thresholds for square-wave electrical pulses. *Science,* 1958, **127**, 879-880.
5. JOHNSON, L. C., & LANDON, M. M. Eccrine sweat gland activity and racial differences in resting skin conductance. *Psychophysiology,* 1965, **1**, 322-329.
6. PLUTCHIK, R. The effect of high intensity intermittent sound on human behavior and physiology. *Psychol. Bull.*, 1959, **56**, 133-151.
7. ———. Auditory pain thresholds for intermittent, "beat" and steady signals. *Percept. & Motor Skills,* 1963, **16**, 863-870.
8. ———. Effect of electrode placement on skin impedance-related measures. *Psychol. Rec.*, 1964, **14**, 145-151.
9. SHAKLEE, A. B. Measurement of cutaneous electrical thresholds in animal research. *Amer. J. Psychol.*, 1957, **70**, 624-627.
10. STERNBACH, R. A., & TURSKY, B. On the psychophysical power function in electric shock. *Psychonomic Sci.*, 1964, **1**, 217-218.
11. VIRGINIA CUTANEOUS PROJECT: 1948-1962. ONR Report NR-140-598, University of Virginia, Charlottesville, 1962.

Department of Psychology
Hofstra University
Hempstead, Long Island
New York 11550

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 155-165.

STRUCTURE OF MATERNAL ATTITUDES IN TWO CULTURES*[1] †

Department of Psychology, The Pennsylvania State University

GEORGE M. GUTHRIE

A. INTRODUCTION

A relationship of at least brief duration between a mother and her child is found in all societies. Probably in most it is prolonged and, unless our theories are wrong, of crucial significance to the development of the personality of the child. This relationship has been the focus of interest in many cross-cultural studies of personality formation in which the antecedent variables are sought in early childhood and the consequent variables in adult behavior and social institutions. As others have pointed out, a number of cultures present a wider range of variations in antecedent conditions than a single culture can offer. Of course, mothers are not the sole socializing agents. There are many societies in which she shares with others responsibility for the care of the child. There is a high likelihood, however, that her attitudes will have much to do with how the child is treated whether she or someone else is with the child at any given moment.

This research is concerned with the degree to which mothers' attitudes, crucial variables in the mother-child relationship, are organized or structured similarly in two societies. Similarity of variables must be assumed or demonstrated if conceptual systems developed in one culture are applied in another. Furthermore, designs which use data from many societies, treating each as an instance, must assume that the same variables are being rated in each culture.

Kardiner (3) has offered a formulation, based on Freudian theory, of the relationship between child-rearing practices and social institutions. Experiences during the oral, anal, and genital stages lay the basis for a society member's projective systems or characteristic modes of reaction which, in turn, determine his belief systems. In much this tradition, Whiting and Child (8), utilizing data from the Human Relations Area Files, examined the relationships between early training experiences and adult beliefs about illness. In research of this sort, the cases are whole societies instead of individuals. By the use of many societies, it is expected that the effects of uncontrolled variables will be randomized and whatever common variance there is between antecedent childhood experiences and adult behavior will become apparent.

---

* Accepted by Leonard W. Doob of the Editorial Board and received in the Editorial Office, Provincetown, Massachusetts, on January 31, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Paper read at annual meeting of the Japanese Psychological Association, July, 1965. The Philippine data were collected during a Fulbright appointment at the Philippine Normal College, Manila.

† This paper has been honored by The Journal Press as a Carl Murchison Memorial Paper.

Minturn and Lambert (4) have examined the antecedents of child rearing in a factor analysis of ratings of 133 mothers from six cultures on a 28-item scale. Because the number of mothers in any given society was small, it was not possible to evaluate the similarity of factors from one society to another. These authors have, however, demonstrated that their orthogonal factors are independent within each of the six societies.

Cross-cultural studies, where correlational procedures are used, must make the assumption that the variables measured vary only in degree from one society to another. Mothers in Society A may be more nurturant than those in Society B, but the nature of nurturance is essentially the same, differing only in frequency or duration in the two societies. The problem of the conceptual equivalence of transcultural variables has been considered by Sears (7), who pointed out that if a personality theory is to be universally applicable, variables are required that can be measured everywhere. However, Whorf (2) has emphasized that different linguistic systems may lead individuals to experience reality differently. If this is so, we may be attaching the same label to patterns of behavior which have markedly different motivational implications.

It may be possible to overcome this problem by using bilingual subjects who permit us to examine, using one language, the relationships between the discrete statements which make up some of the important attitude variables. By this method we can avoid the problem of introducing systematic differences through the incompleteness of translation. Using the same statements of attitudes with *S*s in two different societies may enable us to determine the degree to which behavior is organized so that it is meaningful to use the same summary concepts or variables in the two cultures.

Although the Parental Attitude Research Instrument (PARI) of Schaefer and Bell (6) was developed for use with American mothers, the statements are not too specific to our material culture. The items of this Likert-type questionnaire provide a broad sampling of sentiments that mothers in many societies might hold toward children and family life. Using a heterogeneous population of 413 mothers, Zuckerman *et al.* (9) found three factors, Authoritarian-Control, Hostility-Rejection, and Democratic Attitudes in the 23 scales. Using a somewhat different earlier version of their instrument, Schaefer and Bell (5) found five factors on a population of 100 student nurses. These factors included Suppression of Interpersonal Distance, Hostile Rejection of the Homemaking Role, Excessive Demand for Striving, Over-Possessiveness and Harsh Punitive Control. The first factor is bipolar, and by reversing the direction of interpretation one arrives at something very much like Zuckerman's Democratic Attitudes. While there is considerable agreement concerning the factorial structure of this questionnaire, Becker and Krug (1) found few studies which reported differences on the PARI between groups of mothers where differences were expected. However, the relationships between attitudes and behavior are quite complex. In this research we are concerned with attitudes and not, for the moment, with their complex effects on the children's behavior. Nor was our purpose to determine whether there were differences between the groups on the scales, but to determine by factor analysis the underlying dimensions of the attitudes sampled in the 115 items of the PARI. Our concern was with the comparative organization of attitude systems rather than with the placement of the mothers on continua formulated in one of the cultures involved.

B. PROCEDURE

The Parental Attitude Research Instrument (PARI) of Schaefer and Bell (6) was administered to 133 American and 134 Philippine mothers of elementary-school children. The Americans were members of PTA groups from a rural area in Pennsylvania. The Filipinos were middle-class residents of Manila who were able to take the inventory in English. Although Philippine dialects were spoken in their homes, all of the respondents had attended schools where instruction was entirely in English.

The data from the two groups of *S*s were treated in separate and parallel analyses. The 23 PARI five-item scales were examined for homogeneity of item content. For each of the scales a 5 × 5 correlation matrix was computed and the first factor was extracted. All *S*s were scored for the 23 scales, two 23 × 23 correlation matrices were computed, and six factors were extracted from each matrix. Finally, correlation matrices were computed between items, and 13 factors were extracted from each matrix. Since limitations of our computer restricted us to a 105 × 105 matrix, items from Scales 3 and 6 were not included. The results were interpreted in light of other data available on the two cultures. The number of factors extracted in each case was the same and was approximately the limit for interpretable factors.

C. Results

The analyses of the sets of five items making up the individual scales yielded 23 eigenvalues for each group of mothers. Essentially there were no differences between the two groups of *S*s. Apparently the sets of items making up each scale are such that they correlate reasonably highly within each scale for both groups. It is therefore justified with both groups to add scores on the items to derive scale scores. In terms of these statistical properties, the scales can be used with both samples.

Six factors were extracted from each of the 23 × 23 matrices of correlations between scales. These were rotated, using the varimax solution. The resulting loadings are shown in Table 1.

The factors are listed with tentative names, and with the scales having the highest loadings, in Tables 2 and 3. The first three factors on the American mothers resemble closely those of Zuckerman (9) and Factors V, II and I of Schaefer and Bell (5). Our Factor IV resembles Schaefer's Factor III. In short, four of our factors replicate results obtained by others.

It is clear that the factorial structure is not the same in the two societies. American Factor I emphasizes dominance and feelings of family superiority, while the Filipinos in Factors I and IV treat this domain in terms of loving control and of the importance of the family. Factor II on the American mothers and Philippine Factor III both express discontent. This is accompanied by dependent attitudes in the former, while in the latter it is accompanied by an assertive attitude. American Factor III and Philippine Factor II are very similar. The factors of Spartan Discipline, Female Protest, and Excess Caution have no parallels in the Philippine matrix. Precocity Seeking and Passive Protest are patterns suggested by Philippine data, which were not suggested in the American data.

Thirteen factors were extracted from each of the 105 × 105 matrices of items and rotated, using the varimax solution which retains orthogonality. Tentative names of the 13 factors are presented in Table 4. All factors showed six or more items with loadings above .30. The highest loadings were above .70. The 13 factors accounted for approximately half of the variance in each matrix. The method of analysis did not merely retrieve the five-item scales; although many factors had high loadings from all or most of the items of a scale, it served to isolate some of the principle themes about which mothers

TABLE 1
ROTATED FACTOR LOADINGS OF PARI SCALES FROM PHILIPPINE AND AMERICAN MOTHERS

	Philippine mothers							American mothers						
Scale	I	II	III	IV	V	VI	h^2	I	II	III	IV	V	VI	h^2
1. Encouraging Verbalization	—10	77	—09	17	—02	—11	65	—18	07	—81	21	—10	02	75
2. Fostering Dependency	79	00	02	—14	—05	04	65	70	11	03	28	—03	—20	62
3. Seclusion of the Mother	37	—02	10	—44	—26	—30	50	31	02	—11	—42	34	—42	58
4. Breaking the Will	13	—08	29	—56	—30	—26	59	41	—25	25	—42	30	—28	65
5. Martyrdom	47	—21	20	—26	—44	—18	61	30	—26	08	—03	49	—27	48
6. Fear of Harming the Baby	63	01	31	06	—09	—31	59	32	—30	—27	—13	15	—58	64
7. Marital Conflict	—06	22	75	—19	—14	—05	68	—02	—65	—15	—27	04	13	54
8. Strictness	—04	15	17	—51	01	41	48	08	—07	11	—82	—00	—10	71
9. Irritability	06	—26	78	—12	—04	—14	72	—20	—84	02	—02	—04	02	75
10. Excluding Outside Influences	15	—06	23	—69	06	—01	56	81	05	03	—17	17	—18	75
11. Deification	15	—03	11	—71	—24	03	59	69	05	16	—26	21	—23	62
12. Suppression of Aggression	61	—31	04	—13	—16	—26	59	24	11	10	—10	18	—78	73
13. Rejection of the Home-making Role	28	—34	49	01	—15	—45	65	01	—68	—12	45	13	—11	71
14. Equalitarianism	06	66	16	25	00	25	59	—13	—25	—69	—05	00	—09	56
15. Approval of Activity	02	17	11	—11	—77	—22	69	40	—08	—27	—49	23	—25	59
16. Avoidance of Communication	27	—37	13	—02	—69	—10	71	65	—32	11	—17	03	—13	58
17. Inconsiderateness of Husband	37	—02	—01	—17	—12	—62	56	34	—44	—04	—02	62	—05	71
18. Suppression of Sex	25	—29	—17	—60	—10	—21	59	49	—01	29	—17	36	—42	66

TABLE 1 (*continued*)

Scale	Philippine mothers I	II	III	IV	V	VI	h^2	American mothers I	II	III	IV	V	VI	h^2
19. Ascendance of the Mother	43	08	60	—22	—16	—11	63	05	02	02	—04	86	—14	76
20. Intrusiveness	67	—02	—05	—40	—14	14	66	66	12	02	—18	35	—16	64
21. Comradeship and Sharing	—06	74	—05	—19	—09	—17	62	26	—03	—55	—33	39	30	72
22. Acceleration of Development	13	15	05	—16	—77	02	66	36	07	—10	—41	44	—29	59
23. Dependency of the Mother	—04	09	20	—10	—12	—66	50	38	—52	—03	—07	27	—22	54

TABLE 2
FACTORS FOR PHILIPPINE MOTHERS FROM PARI SCALES

Factor	Scale	Loading
I Loving Control	2. Fostering Dependency	.79
	20. Intrusiveness	.67
	6. Fear of Harming Baby	.63
	12. Suppression of Aggression	.61
II Democratic Attitudes	1. Encouraging Verbalizations	.77
	21. Comradeship and Sharing	.74
	14. Equalitarianism	.66
III Resentment	9. Irritability	.78
	7. Marital Conflict	.75
	19. Ascendance of the Mother	.60
IV Importance of Family	11. Deification	.71
	10. Excluding Outside Influences	.69
	18. Suppression of Sex	.60
	4. Breaking the Will	.56
	8. Strictness	.51
V Precocity Seeking	22. Acceleration of Development	.77
	15. Approval of Activity	.77
	16. Avoidance of Communication	.69
VI Passive Protest	23. Dependency of Mother	.66
	17. Inconsiderateness of Husband	.62
	13. Rejection of Homemaking Role	.45

in the two societies organized their attitudes about parent-child relationships. The tentative names of the factors are based, of course, on an examination of the content of the items with high loadings.

The outstanding difference between the two samples lies in the number of factors related to parental authority and control. While in the American sample three factors (A, F, and M) deal with parental control, the Philippine data show seven factors (A, C, E, G, H, J, and K) related to the parent's control of her child.

In contrast, five factors (C, D, H, J, and L) in the American data relate to the child's being on his own and responsible for his own behavior. The Philippine data yield two such factors, F and M. These findings are consistent with data obtained by observation that there is a good deal more attention devoted to parental control and influence in the Philippines, while American mothers are more concerned with establishing their child's independence. The domain of parent-child relationships would seem therefore to be organized differently in the two societies.

Using the data from the analyses of the scales and of the items, we can get some indication whether the same behavior patterns covary in the two

TABLE 3
FACTORS FOR AMERICAN MOTHERS FROM PARI SCALES

Factor	Scale	Loading
I Dominance, Feeling of Family Superiority	10. Excluding Outside Influences	.81
	2. Fostering Dependency	.70
	11. Deification	.69
	20. Intrusiveness	.66
	16. Avoidance of Communication	.65
II Resentment, Weakness	9. Irritability	.84
	13. Rejection of Homemaking Role	.68
	7. Marital Conflict	.65
	23. Dependency of Mother	.52
III Democratic Attitudes	1. Encouraging Verbalization	.80
	14. Equalitarianism	.69
	21. Comradeship and Sharing	.55
IV Spartan Discipline	8. Strictness	.82
	15. Approval of Activity	.49
	13. Rejection of Homemaking Role	.45
	3. Seclusion of Mother	.42
	4. Breaking the Will	.42
V Female Protest	19. Ascendance of Mother	.86
	17. Inconsiderateness of Husband	.62
	5. Martyrdom	.49
	22. Acceleration of Development	.44
VI Reaction Formation of Excess Caution	12. Suppression of Aggression	.78
	6. Fear of Harming Baby	.58
	18. Suppression of Sex	.42
	3. Seclusion of Mother	.42

TABLE 4
SUGGESTED NAMES OF FACTORS EXTRACTED FROM RESPECTIVE 105 × 105 MATRICES OF PARI ITEMS

American	Philippine
A. Parental supremacy	A. Authority of parent
B. Maternal irritability	B. Comradeship and fun
C. Spartan discipline	C. Fostering closeness of family
D. Tolerance of differences	D. Woman's strong role
E. Ascendance of mother	E. Strong control
F. Fostering dependency	F. Acceleration of development
G. Isolation of the mother	G. Controls for respect
H. Social decorum	H. Parental authority
I. Martyrdom of the mother	I. Loss of freedom of mother
J. Comradeship	J. Parental responsibility
K. Intrusiveness	K. Parental wisdom
L. Fostering independence	L. Maternal role of strictness
M. Disinterested control	M. Equality to children

societies. From these patterns of covariation we can infer the degree of transcultural equivalence of some of the important variables in parent-child relationships. We can interpret the meaning of an item or group of items by examining not only the content of the item or group, but by paying attention also to the items that correlate most highly with those with which we are concerned. For example, Scale 4, Breaking the Will, in the American sample is part of a cluster of behavior which emphasizes a Spartan approach to life, a determined attempt to develop character through self-sacrifice. The same items, which form a satisfactory scale on Philippine mothers, correlate with items which reiterate the importance of family considerations. In short, the same items referring to the same attitudes and behavior have quite different meaning and are probably motivated by different purposes.

Scale 2, Fostering Dependency, in the Philippine sample is part of a constellation of behavior related to family support and family responsibility for the child, while with our American mothers it is related to maternal control and authority. Scale 5, Martyrdom, appears as an expression of protest, while on Philippine mothers the scale loads factors suggesting loving control and promotion of development of the child. Fear of Harming the Baby was associated with other avoidance activities in American mothers, while in the other group it appeared on a factor of loving control. For Filipinos, Strictness was associated with family reputation, while American mothers saw it a matter of disciplined development. Irritability was associated with assertiveness in the Philippines and with dependence in the United States. Suppression of Sex interests was largely a matter of retaining the importance of the Philippine family, while the American mother linked this attitude with both family control and a system of avoidances. Employing this technique of analysis makes possible the identification of differences in the meaning and emotional significance of many of the 23 variables assessed by the set of items used.

D. Discussion

The use of an inventory without translation on mothers from two quite different societies has permitted the comparison of the structure of mothers' attitudes toward their children. Since the individual scales showed similar eigenvalues, and the other pairs of matrices showed approximately similar communalities, we can conclude that the reliabilities in the two samples were essentially the same. We can therefore ask whether the variables, as defined by the reports of the mothers, are the same in the two cultures.

Our data suggest that much caution must be observed in using variables defined in one culture in the study of personality development in another. It would appear that a pattern of behavior encompassed by given items may be homogeneous, but the whole pattern may occur in a different context of motivation and therefore have quite different significance in one society than in another. The same set of items is not measuring the same thing in each instance.

We believe that what we have found is that mothers in two societies may say, or do, the same things, but the meaning of these superficially similar activities may be quite different. The disparity becomes most apparent when we examine the patterning of beliefs or behavior within which the discrete acts in which we are interested occur. If this is so, cross-cultural comparisons are made more difficult because we cannot assume the same continua in the cultures under comparison.

Psychodynamic theories, all profoundly influenced by Freud, have been formulated within a more or less restricted cultural context. Their biological and instinctual biases have implied a more universal validity than there are data to support. While many non-Western scientists have found the concepts useful, our data suggest that there are important differences in such variables as dependency, negativism, dominance, etc. Insofar as cultures differ it may be necessary to develop somewhat different formulations of the relationships between antecedent experiences and consequences in the behavior of the individual.

By way of summary, factor analyses of PARI data on Philippine and American mothers suggested differences in the structure of maternal attitudes. Although the individual five-item scales had comparable eigenvalues, the factorial structure of the matrix indicated that superficially similar behavior had different meanings in the two societies. These results suggest caution in assuming similarity of behavior variables in cross-cultural research.

REFERENCES

1. BECKER, W. C., & KRUG, R. S. The parent attitude research instrument—A research review. *Child Devel.,* 1965, **36**, 329-365.
2. CARROLL, J. B., *Ed.* Language, Thought and Reality: Selected Writings of B. L. Whorf. Cambridge, Mass.: Technology Press, 1956.
3. KARDINER, A. Psychological Frontiers of Society. New York: Columbia Univ. Press, 1945.
4. MINTURN, L. G., & LAMBERT, W. W. Mothers of Six Cultures. New York: Wiley, 1964.
5. SCHAEFER, E. S., & BELL, R. Q. Patterns of attitudes toward child rearing and the family. *J. Abn. & Soc. Psychol.,* 1957, **54**, 391-395.
6. ———. Development of a parental attitude research instrument. *Child Devel.,* 1958, **29**, 339-361.

7. SEARS, R. R. Transcultural variables and conceptual equivalence. In B. Kaplan (Ed.), *Studying Personality Cross-Culturally*. Evanston, Ill.: Row, Peterson, 1961. Pp. 445-455.

8. WHITING, J. M. W., & CHILD, I. Child Training and Personality. New Haven, Conn.: Yale Univ. Press, 1953.

9. ZUCKERMAN, M., RIBBACK, B. B., MONASHKIN, L., & NORTON, J. A., JR. Normative data and factor analysis on the parental attitude research instrument. *J. Consult. Psychol.*, 1958, **22**, 165-171.

Department of Psychology
The Pennsylvania State University
117 Burrowes Building
University Park, Pennsylvania 16802

Published as a separate and in *The Journal of Psychology*, 1966, 62, 167-175.

THE ANATOMY OF ABNORMALITY*

St. John's University; and Palo Alto Mental Research Institute

William G. Herron and Robert E. Kantor

A. Introduction

Scrutiny of the concept of abnormality indicates ferment and need for change. Abnormality appears to be a clear, logical concept when the norm is the average, common occurrence and abnormality refers to infrequent events and unusual behavior. If the possibility of clearly specifying the behavior to be studied is assumed, abnormality can then be objectively measured.

However, the statistical concept of abnormality is not successful in reducing the connotative aspects of the term. Thus, those whose behavior places them above the norm will usually be called superior rather than abnormal. This is due to the negative connotations attached to being abnormal. The culture places an affective charge on the concept which in turn delimits its use.

In fact, abnormality is normally equated with undesirable, maladaptive, pathological, disorganized, maladjusted states classified as mental illnesses and mental diseases. The equation necessitates a value judgment of the behavior under consideration. The value norm is often not statistically determined, nor many times could it be, for personality disorganization and its equivalent are more notable for their existence than their adaptability to either specification or quantification.

The major task of defining the behavior to be considered as abnormal can be illustrated by comparing four definitions of schizophrenia available from a much larger pool of definitions of what is supposed to be an entity.

". . . a group of psychotic reactions characterized by fundamental disturbances in reality relationships and concept formations, with affective behavioral, and intellectual disturbances in varying degrees and mixtures. The disorders are marked by a strong tendency to retreat from reality, by emotional disharmony, unpredictable disturbances in stream of thought, regressive behavior, and in some, by a tendency to 'deterioration'" (1, p. 26).

"Schizophrenia is, in its acute form, an intense fear reaction, presented to the observer as either a panic or a state of catatonia which must be considered

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 2, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

as equivalent of a panic. In its chronic forms it is seen as various reactions, usually designated as paranoid hebephrenic or as 'social recovery'" (45, pp. 180-181).

"Schizophrenia is a specific reaction to an extreme state of anxiety, originating in childhood, and reactivated later in life by psychological factors. The specific reaction consists of the adoption of archaic mental mechanisms, which belong to lower levels of integration" (2, p. 384).

"The disease is characterized by a specific type of thinking, feeling, and relation to the external world. . . . Certain symptoms of schizophrenia are present in every case. . . . Beside these specific permanent or fundamental symptoms, we can find a host of other, more accessory manifestations. . . . The fundamental symptoms consist of disturbances of association and affectivity, the predilection for fantasy as against reality, and the inclination to divorce oneself from reality (autism)" (5, pp. 13-14).

Other definitions stressing more divergent viewpoints could be added to underscore the evasive quality of schizophrenia. While rapprochement may be possible in regard to defining schizophrenia, the complexity of the concept would continue to be apparent; and the breakdown of any of the operations involved in schizophrenia into measurable terms is not comparable to the statistical norms that can be obtained for a concept such as scholastic achievement.

The concept of abnormality is essentially stretched to cover a multitude of behaviors often lacking common elements and eluding both satisfactory identification and measurement. An awareness of the problem is seen in a current trend in abnormal psychology to recognize in detail the shortcomings of the various meanings that can be attributed to abnormality and, from the inadequate, to attempt the selection of some definition that represents a certain adequacy. Thus, abnormality is represented by social maladjustment (7); personality disorganization (26); unusual, unadaptable, and pathological behavior (29); personal and social maladjustment (40); and disordered personal reactions (44).

These terms are connotative compromises between calling abnormality mental disease and using designations such as problems in living (39) or sin (32). The compromise reflects the fact that the behavior called abnormal cannot be designated as illness or disease without considerable translation (18, 27). Menninger refers to current terminology as "obsolete handles which have been so long used, arranged, ordered, rearranged, reordered, disputed, and at last, we hope, discarded" (31, p. 48).

B. ABNORMALITY AND THE CULTURE

In a changing conception of abnormality, recognition should be given to the relative importance of the culture and the subculture in regard to the behavior that is judged abnormal. This goes beyond the awareness that abnormal behavior in one setting is normal in another and involves emphasizing the values of the environment rather than the behavior of the individual in the environment; for the determination of abnormality can rest more upon the shifting value judgments of society than upon the commission of specific acts.

Eysenck (11) has suggested a paradigm of research on abnormality in which a change in the independent variable effects a change in a function making it more or less abnormal. While this is intended as a method of studying the functioning of psychological systems of individuals, it also leaves the way open for an emphasis on the behavioral environment. This emphasis is clearly suggested in Ferster's (13) description of abnormality as the result of differential reinforcement, and as comprehensible in terms of the mutually reinforcing behavior of the subject and his environment. "Whether a man who moves and acts slowly is 'depressed' or merely moving slowly is not easily or reliably determined by observing his behavior alone. The relation of his behavior to events in the past or present environment is a critical element in the description" (13, p. 9).

Abnormality needs to be viewed as an interactive process. Its determination involves studying the context of behavior as well as the behavior. For example, the subjects comprising abnormal samples for research purposes are very frequently drawn from locations containing "known" abnormals. The certainty of the abnormality is supposedly determined by the behavior of the subjects, but it is also determined by the fact that the subjects are at a place designated to be for abnormal people. Because of where they are, their abnormality is underscored by them and by their environment; and to be there is to be abnormal in a consequential manner ranging from the cost and effort of outpatient counseling to the extremes suggested by Orlans (33) in comparing the mental hospital to a concentration camp.

The role of the culture is plain in the case of the covert abnormals. These are people who are not at a place for the abnormal, although they could be. Their grossly deviant behavior is often not noticed or, if it is noticed, it is rationalized as eccentricity. Many persons who "fantasy" events, who are unusually withdrawn, who hear voices or see visions, are not necessarily labeled as insane by others (8).

That is to say, the *total* prevalence of mental disorder in a community is much higher than the *treated* prevalence (37). In a Baltimore study (34), the rate of treatment in VA, state, and private hospitals was .5 per cent, as compared with the untreated rate of 10.9 per cent. Since patients treated by private physicians were not included, probably the .5 per cent figure is too low. The figure of .8 per cent is reported for New Haven (20). In an investigation conducted at Kalamazoo, Michigan (30), which carefully arrived at a very conservative estimate, the discrepancy still prevailed between total and treated prevalence. Of a population of 164,125 persons, .6 per cent were in psychiatric care, whereas 3.4 per cent were identified as significantly ill.

These data suggest that it might be useful to divide the study of the etiology of abnormality into different parts. First, how do we explain the psychological processes that produce mental deviation in our population? This is the usual question, but it is traditionally asked only about the treated sample, the .5-.8 per cent. The second part of the problem is neglected as yet: among the total population of deviants, what are the ways in which only a small portion is selected for psychiatric attention?

C. Symptoms

Mental health disorders are supposed to be translatable in terms of symptoms, but symptomological descriptions of abnormality merit the criticism they have gotten. The need is for a meaningful conception of the dangers and hurts in everyday existence.

Schizophrenia is an apt example of the symptom problem. It is a label attached with such uncertain frequency that the temptation is strong to eliminate the term and its cluster of symptoms. Yet, if schizophrenics are grouped into those who seem to recover and those who do not, the symptom pictures become relatively distinct (22). For example, nonrecoverable patients are characterized by primitive behavior, such as withdrawal and reality distortion (4, 28). In accord with this is the finding that the symptoms of the recoverable patient are developmentally higher than the nonrecoverable's symptoms (46).

In fact, much of the diverse and apparently unrelated symptomatic behavior of schizophrenics can be meaningfully ordered by using the process-reactive classification of schizophrenia (17, 19, 21, 36). Schizophrenia is described as a continuum of personality disorganization ranging from the greatest disintegration at the process pole to the threshold of normality at the reactive end. The typical process syndrome is characterized by perceptual inefficiency and immaturity, inadequate premorbid social and sexual functioning, inappropriate interpersonal relationships, and cognitive malfunctioning. In contrast, the typical reactive syndrome is a kind of "reasonable break with reality," in which psychological functioning remains relatively mature and the prognosis is good (22).

Of course, these are typical rather than universal patterns. For example, Herron (16) found significant differences between process and reactive patients on tests of abstract ability, while Becker (3) found differences for males but not for females, and Eliseo (9, 10) failed to find differences. While there are specific variations in these studies, as in the tests used to measure abstract ability, their significance lies in pointing to the atypicality of specific behavior. The patients are not psychotic all the time.

Symptoms can be meaningfully grouped as typical, but not universal, patterns of behavior related to the recovery potential of the patients. The first task is the identification of the typical patterns. This is currently being attempted in regard to schizophrenia, but certainly merits extension to all syndromes of mental illness. The second task is the identification of the atypical symptom patterns and their meaning, since even symptoms which appear integral to a specific disorder will not always appear in a patient's behavior. The varying perceptions of stress by different patients affect the symptom picture. Some symptoms may be "picked up" almost accidentally. Symptoms useful at one time are discarded at another in response to the whims of the environment.

Symptom descriptions refer to the person at a particular moment in time. Repeated observations will show certain consistencies and inconsistencies. The consistent symptoms will be more easily observable and more obviously related to developmental events and recovery potentials than will the inconsistent symptoms, but in neither case will symptoms be immutable or always observable.

Symptoms have an enormous relativity which merits emphasis in any conceptual system of impaired psychological functioning. For example, the reactive schizophrenic has a notoriously normal appearance (16, 38), except that at some time he behaved in such a way as to be designated most abnormal. The process schizophrenic looks very ill, except for his normal moments. The picture is completed by the psychopathology of normal living. It seems almost unfriendly to talk about normal symptoms, since symptoms are considered signs of mental disorder, and the border of illogicality may appear to be touched with the mention of the healthy schizophrenic. Nevertheless, what often becomes apparent is the abnormality of normality, and the normality of abnormality.

Despite these complications, there are some clear findings that differentiate reactive schizophrenics from process ones. One of these findings is that reactive schizophrenics usually have held a job in their society. When they recover from their collapses, they come back to the job. When the job is not available, the reactive schizophrenic may deteriorate into a process state. The picture is different for the process schizophrenic. It is characteristic that he began to show failure much earlier in life. Usually even in school play and in school lessons he is reported as inadequate. The jobs he holds in adulthood are generally very low in calibre and are transitory.

Possibly the employment situation clarifies why only a minority of schizophrenics are selected for institutionalization and why this minority comes mostly from the least privileged social classes (20). In those classes, worklessness is a serious transgression. The process schizophrenic shows work incapacity very soon in life, while the reactive schizophrenic works later into life. Each is punished by society in accordance with his work history. Typically, the process schizophrenic may be institutionalized for much longer periods than is the reactive schizophrenic.

D. Prognosis

Whether or not a patient would recover has often been viewed as dependent on his specific disorder. One result of such an approach has been the fatalism adhering to the label of schizophrenia. Yet, within the diagnostic grouping called schizophrenia, there are degrees of personality decompensation and corresponding differences in recovery potentials (18, 19, 22, 42, 43). The indications are that the potential for recovery depends on the degree of personality maturity in the mentally ill person. This can be demonstrated by research on prognosis in schizophrenia.

Process-reactive criteria were initially developed on the basis of their occurrence in developmental periods (25). Subsequent research showed that severity of schizophrenia was related to the developmental epoch involving severe pathogenic events (23, 24). Schizophrenics vary in their degree of pathology according to their degree of premorbid social and sexual maturity (14, 35, 46). Patients varying in prognosis also show differences in their self-perception, with the reactive patient describing himself as more mature than does the process patient (17, 41).

These results place mental illness in a multidimensional personality range. Within this range are certain relatively identifiable integrative and disintegrative patterns and an as yet uncounted number of atypical patterns. Hopefully, there is always a health-embryo-within-illness, though its birth rate would be intimately dependent on individual developmental maturity.

Also, while the reactive patient has considerable recovery potential, society may limit its actualization. For example, the structure of many a mental hospital may make the healthier patient regress (22). This regression is easily seen in the peculiar hospital practice of "identity stripping" (15). In this practice, the patient loses his personal property, his bodily privacy is violated in the form of medical examinations, his psychological functions are intimately probed with an emphasis on hidden deviations, and he is constantly watched and regimented. Most important, he undergoes powerful pressures towards the acceptance of a self-image as one who is sick, dependent, and irresponsible. Regression, or the appearance of it, can be seen as a natural expression of rejection of the mental hospital.

"When a patient finds himself in seclusion, naked and without visible means of expression, he may have to rely on tearing up his mattress, if he can, or writing with feces on the wall—actions management takes to be in keeping with the kind of person who warrants seclusion" (15, pp. 306-307).

In contrast to this dismal situation, some recent studies have held out hope of adjustment even for process schizophrenics, when jobs could be offered to them. Fairwether (12) compared 169 schizophrenics on the dimensions of duration of hospital residence and of many different kinds of therapy provided to these patients. His finding was that treatment, personnel attitudes, and patient aspirations were all irrelevant to posthospital success. The factor basic to the patient's success was a "supportive living situation accompanied by employment" (12, p. 278). A British study (6) following up the adjustive success of 76 schizophrenic patients, each rated as a poor risk by psychiatric standards, revealed that 40 were definitely successful on the outside. The successful ones were those who held a job; the unsuccessful ones remained unemployed.

E. Conclusion

Abnormality is a concept in transition that is most often equated with behavior that is difficult to identify and to measure. The connotative aspect of the concept is enormous and plays a great part in the difficulties involved in trying to understand what is meant by abnormality. There is a need for reformulating abnormality so that it can be viewed as an interactive construct. Significant account should be taken of the range of behavioral, etiological, and prognostic differences within the currently designated abnormal syndromes.

REFERENCES

1. AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and Statistical Manual. Mental Disorders. Washington: American Psychiatric Association, 1952.
2. ARIETI, S. Interpretation of Schizophrenia. New York: Brunner, 1955.
3. BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. Abn. & Soc. Psychol.*, 1956, **53**, 229-236.
4. ———. The process-reactive distinction: A key to the problem of schizophrenia. *J. Nerv. & Ment. Dis.*, 1959, **129**, 442-449.
5. BLEULER, E. Dementia Praecox or the Group of Schizophrenias. New York: Internat. Univ. Press, 1950.
6. CARSTAIRS, G. M. The social limits of eccentricity: An English study. In M. K. Opler (Ed.), *Culture and Mental Health.* New York: Macmillan, 1959. Pp. 373-389.
7. COLEMAN, J. C. Abnormal Psychology and Modern Life. Fairlawn: Scott, Foresman, 1956.
8. CUMMING, E., & CUMMING, J. Closed Ranks. Cambridge: Harvard Univ. Press, 1957. Pp. 92-103.
9. ELISEO, T. S. Figurative and literal misinterpretations of words by process and reactive schizophrenics. *Psychol. Rep.*, 1963, **13**, 871-877.
10. ———. Overinclusive thinking in process and reactive schizophrenics. *J. Consult. Psychol.*, 1963, **27**, 447-449.
11. EYSENCK, H. J. Classification and the problem of diagnosis. In H. J. Eysenck (Ed.), *Handbook of Abnormal Psychology.* New York: Basic Books, 1961. Pp. 1-31.
12. FAIRWETHER, G. Social Psychology in Treating Mental Illness. New York: Wiley, 1964.
13. FERSTER, C. B. Classification of behavioral pathology. In L. Krasner & L. P. Ullmann (Eds.), *Research in Behavior Modification.* New York: Holt, Rinehart & Winston, 1965. Pp. 6-26.
14. GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *J. Nerv. & Ment. Dis.*, 1959, **129**, 450-466.
15. GOFFMAN, E. Asylums. New York: Anchor, 1961.
16. HERRON, W. G. Abstract ability in the process-reactive classification of schizophrenia. *J. Gen. Psychol.*, 1962, **67**, 147-154.
17. ———. The process-reactive classification of schizophrenia. *Psychol. Bull.*, 1962, **59**, 329-343.
18. ———. The disease controversy: Closing open issues. *Psychol. Rep.*, 1963, **13**, 79-84.
19. HIGGINS, J. The concept of process-reactive schizophrenia: Criteria and related research. *J. Nerv. & Ment. Dis.*, 1964, **138**, 9-25.
20. HOLLINGSHEAD, A. B., & REDLICH, F. C. Social Class and Mental Illness. New York: Wiley, 1958.
21. JOHANNSEN, W. J. On classifying schizophrenics: A modest proposal. *Psychol. Rep.*, 1964, **14**, 38.
22. KANTOR, R. E., & HERRON, W. G. Reactive and Process Schizophrenia. Palo Alto: Sci. & Behav. Books, in press.
23. KANTOR, R. E., & WINDER, C. L. The process-reactive continuum: A theoretical proposal. *J. Nerv. & Ment. Dis.*, 1959, **129**, 429-434.

24. ———. Schizophrenia: Correlates of life history. *J. Nerv. & Ment. Dis.*, 1961, **132**, 221-226.

25. Kantor, R. E., Wallner, J. M., & Winder, C. L. Process and reactive schizophrenia. *J. Consult. Psychol.*, 1953, **17**, 157-162.

26. Kisker, G. W. The Disorganized Personality. New York: McGraw-Hill, 1964.

27. Linder, R. Diagnosis: Description or prescription? A case study in the psychology of diagnosis. *Percept. & Motor Skills*, 1965, **20**, 1081-1092.

28. Lorr, M., Wittman, P., & Schanberger, W. An analysis of the Elgin Prognostic Scale. *J. Clin. Psychol.*, 1951, **7**, 260-263.

29. Lundin, R. W. Principles of Psychopathology. Columbus: Merrill, 1965.

30. Manis, J. G., Brawer, M. J., Hunt, C. L., & Kercher, L. G. Estimating the prevalence of mental illness. *Amer. Sociol. Rev.*, 1964, **29**, 84-89.

31. Menninger, K. The Vital Balance. New York: Viking, 1963.

32. Mowrer, O. H. The Crisis in Psychiatry and Religion. New York: Van Nostrand, 1961.

33. Orlans, H. An American death camp. In B. Rosenberg, I. Gerner, & F. W. Howton (Eds.), *Mass Society in Crisis.* New York: Macmillan, 1964. Pp. 614-628.

34. Pasamanick, B. A survey of mental disease in an urban population: IV. An approach to total prevalence rates. *Arch. Gen. Psychiat.*, 1961, **5**, 151-155.

35. Phillips, L. Case history data and prognosis in schizophrenia. *J. Nerv. & Ment. Dis.*, 1953, **117**, 515-525.

36. Rabin, A. I., & King, G. F. Psychological studies. In L. Bellak (Ed.), *Schizophrenia: A Review of the Syndrome.* New York: Logos Press, 1958. Pp. 216-278.

37. Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., & Rennie, T. A. C. Mental Health in the Metropolis: The Midtown Study, Vol. I. New York: McGraw-Hill, 1962.

38. Strum, I. E. Overinclusion and concreteness among pathological groups. *J. Consult. Psychol.*, 1965, **29**, 9-18.

39. Szasz, T. S. The Myth of Mental Illness. New York: Hoeber-Harper, 1961.

40. Thorpe, L. P., Katz, B., & Lewis, R. T. The Psychology of Abnormal Behavior. New York: Ronald, 1961.

41. Ullmann, L. P., & Giovannoni, J. M. The development of a self-report measure of the process-reactive continuum. *J. Nerv. & Ment. Dis.*, 1964, **138**, 38-42.

42. Vaillant, G. E. The prediction of recovery in schizophrenia. *J. Nerv. & Ment. Dis.*, 1962, **135**, 534-543.

43. ———. An historical review of the remitting schizophrenias. *J. Nerv. & Ment. Dis.*, 1964, **138**, 48-56.

44. White, R. The Abnormal Personality. New York: Ronald, 1964.

45. Woolley, L. F. Experiential factors essential to the development of schizophrenia. In P. H. Hoch & J. Zubin (Eds.), *Current Problems in Psychiatric Diagnosis.* New York: Grune & Stratton, 1953. Pp. 180-191.

46. Zigler, E., & Phillips, L. Social competence and the process-reactive distinction in psychopathology. *J. Abn. & Soc. Psychol.*, 1962, **65**, 215-222.

Department of Psychology
St. John's University
Grand Central and Utopia Parkways
Jamaica, New York 11432

Mental Research Institute
777 Bryant Street
Palo Alto, California

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 177-178.

A NOTE ON SMART'S STUDY OF BIRTH RANK AND AFFILIATION IN MALE UNIVERSITY STUDENTS*

Addiction Research Foundation, Toronto, Ontario, Canada

JAN E. E. DE LINT

A recent study by Smart (2) reports that first and only born male students are more frequently members of social and recreational clubs than are later born male students. This finding is taken to support the argument that affiliative behavior in males varies with birth rank.

However, it is quite conceivable that the difference found by Smart was not due to differential affiliative behavior by rank of birth, but to a class bias introduced when first and only borns were compared to later borns. Lower-class families tend to be larger and hence to contain more later borns than do upper-class families. Since lower-class persons are probably also less likely to belong to recreational and social clubs (for both social and economic reasons), the class variable should have been controlled in a comparison of this sort.

One way to control for social class (and for all other factors that may be associated with sib size) is to consider first and last borns only. Any sib, regardless of its size, has only one first and one last born (1). Hence any population, and therefore any social-class group, must contain equal numbers of these. Therefore, if affiliative behavior varies with birth rank in males, and if club membership can be taken to indicate such behavior, then more first born than last born males should be club members.

To test this possibility, the self-administered questionnaires used by Smart[1] were re-examined; the relevant data are summarized in Table 1.

Since there is no difference in the frequency of first and last born male students with reference to club membership ($\chi^2 = 0.3$; $p > .50$), it follows that the difference found by Smart between first and later borns is attributable to the relatively large number of "other borns" falling into his "later borns" category. In other words, the latter category tends to contain more persons from large families and, therefore, more from low socioeconomic strata than

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 4, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] These questionnaires were collected by Smart and the author as part of a larger study of smoking habits.

TABLE 1
BIRTH RANK AND CLUB MEMBERSHIP IN MALE UNIVERSITY STUDENTS
(Expected frequencies in parentheses)

Membership in social and/or recreational clubs	Birth rank		Total
	First	Last	
Belong	46 (44.4)	33 (34.4)	79
Do not belong	20 (21.4)	18 (16.6)	38
Total	66	51	117

the former category contains. In sum, if family size—and therefore social class—is controlled, and if club membership can be taken to measure affiliative behavior in males, no association between such behavior and birth rank is obtained.

REFERENCES

1. GREGORY, I. An analysis of familial data on psychiatric patients: Parental age, family size, birth order and ordinal position. *Brit. J. Prev. & Soc. Med.*, 1958, **12**, 42-59.
2. SMART, R. G. Social-group membership, leadership and birth order. *J. Soc. Psychol.*, 1965, **67**, 221-225.

Alcoholism & Drug Addiction Research Foundation
24 Harbord Street
Toronto 5, Ontario, Canada

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 179-194.

PSYCHOLOGICAL NEEDS AND VERBALLY EXPRESSED AGGRESSION OF ADOLESCENT DELINQUENT BOYS*

Departments of Psychology, Ohio State University and University of Minnesota

JOHN E. HORROCKS AND NATHAN W. GOTTFRIED

A. INTRODUCTION

This study examines certain differences between delinquent and nondelinquent adolescents in the directional aspects of behavior. Throughout the literature on adolescence the hypothesis has been proposed (2, 3, 10) that delinquent behavior is an aggressive response pattern related to frustration in the attainment of such relatively primary goals as affection, security, and recognition. Writing on this topic, Horrocks (4, p. 494) notes that "an adolescent . . . who is frustrated because his goals are blocked will tend to display aggressive behavior, the amount and strength of the aggressive behavior being a direct function of the frustration."

The existence of a possible relationship between delinquency and frustration-aggression having been assumed, two hypotheses are considered for the present study: (*a*) delinquents and nondelinquents differ in goal attainment, and (*b*) delinquents and nondelinquents differ in aggression. Goal attainment is held to result from the existence of an hierarchy of needs, the theoretical basis of the structure "need" being that used by Lucas and Horrocks (8), Benimoff and Horrocks (1), and by others during a series of investigations of adolescent behavior in a midwestern city.

B. PROCEDURE

The two hypotheses proposed for this study were tested by two different instruments using two different samples of subjects. An adolescent needs inventory devised by Lucas, Horrocks, and others (9) was used to determine the need-satisfaction status of delinquents as compared to nondelinquents. Use of this instrument permitted two types of analyses of the need pattern of delinquents: (*a*) classification of needs in terms of psychological variables, and (*b*) classification of needs on the basis of "behavioral situations" or environmental settings.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 4, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

To determine whether or not adolescents differ in "aggression" as a typical mode of response, thematic picture stories consisting of four TAT pictures and six pictures devised for this study were used.

Several factors were considered in the decision to use the thematic pictures:

1. It was believed that the instrument might discriminate between delinquent and nondelinquent stories on the ratings of "aggression."
2. A scoring system based on the content of stories could be used instead of inferences about "underlying traits" about which there appears to be much disagreement.
3. The ratings could be treated statistically.
4. The raw data (tape recordings) could be preserved for future research.

C. SUBJECTS

All subjects utilized in this sample were school-age children living in a southwestern Ohio city of 250,000 persons. Sample I to whom the needs inventory was administered consisted of 27 boys judged as delinquents by the courts and held at a county detention home, and of 55 male ninth-grade students without records of delinquent behavior. The detention-home boys, all of whom volunteered to participate in the study, constituted the "delinquent" group, while the ninth graders all enrolled in a local high school constituted the comparison ("nondelinquent") group. The high school utilized was chosen from among the 10 high schools located in the city school district because its student body was drawn largely from an area characterized by low occupational and housing levels (U.S. Census of Population: Vol. III, Census Tract Statistics, 1950). Ages of subjects in Sample I ranged from 12-17 years with a median age of 14.4 for the delinquent group and from 13-17 years with a median age of 14.7 for the control group. The median *IQ* was 86.5 for the delinquent and 90.0 for the comparison group.

Sample II to whom the thematic picture test was administered consisted of 21 male delinquent-nondelinquent pairs matched for age, intelligence, and race. The 21 pairs were drawn from two subsamples. Subsample 1 consisted of 50 delinquents of record attending high school in the city district. Their names were obtained through the Juvenile Court, the Pupil Personnel Department of the city schools, and from high-school guidance counselors. In each case, the subject's record in the Pupil Personnel file was checked to confirm that he had actually been adjudged delinquent by the Juvenile Court. Due to factors of mobility, truancy, and of the "elusiveness" of some of the individuals, it was possible to include only 21 of the original 50 in the study.

Nondelinquents for Sample II were selected from a random subsample of

153 high-school and elementary-school children, without delinquency records, attending the same schools as the delinquent subsample described above. From this subsample of 153, 21 individuals were selected on the basis of age, intelligence, and race to pair with the 21 delinquents already selected. The adequacy of the matching procedure may be judged by the fact that no pair differed by more than one year in age nor by more than 6 *IQ* points. Median age of the subjects in Sample II was 15.2 years (range 12-17 years); median *IQ* was 90 (range 70-115). Eleven pairs were Negro and 10 pairs were whites. It was not possible to match for socioeconomic status on an individual basis, but partial matching was accomplished by restricting both the delinquent and the nondelinquent samples to a single high-school district in a lower socioeconomic section of the city. Matching on the basis of race also tended to minimize differences in socioeconomic status.

D. INSTRUMENTS

1. *The Needs Inventory*

The needs inventory was constructed by Lucas and others (9) to identify the "needs" of adolescents with reference to categories and definitions utilized in a series of psychological investigations of the adolescent population of the city of Delaware, Ohio, conducted by Horrocks and his associates.

In a survey of the literature on adolescence, Lucas (6, 7) found frequent use of the term "need," but little agreement as to its definition. Four major types of investigations were used by various investigators to infer need categories: behavioral observations, case studies, studies of free responses of adolescents to broad questions, and studies of response to problem checklists or inventories. However, differing theoretical approaches and methodology resulted in a multitude of terms describing what appeared to be similar concepts pertaining to adolescent motivation, and a few terms that seemed to have widely different meanings.

To order the terminology and definitions applied to adolescent "needs," Lucas compiled a list of "needs" of adolescent youth postulated in the literature. The terms were categorized by seven judges (advanced graduate students and senior staff members, Department of Psychology, Ohio State University) by sorting and resorting until agreement on categories was reached. The criteria used in the sorting were as follows: (*a*) Each need must be relatively independent of other needs in the system. (*b*) Each need must be independent of nonneed variables. (*c*) The needs themselves must be neither too broad nor too specific abstractions of the behavior referents.

Guiding the sorting team was the following definition to the term "need" formulated by Lucas:

> an abstraction from a classification of the observed and directional aspects of behavior. This definition involves as few assumptions as possible regarding the nature and operation of the construct. The definition of need, as it is to be understood in the present study, requires but one assumption—behavior has observable directionality. The observable can be classified along qualitative and quantitative lines with reference to its directionality. To observe behavior, to distinguish therein a particular pattern and then to infer from it a hypothetical construct called "need" is logically sound. The inference of "needs" from observed behavior is actually a method of classifying the observable. "Need" here is not defined as an entity; it is not a "thing" independent of the operations which define it. This definition avoids the innate versus acquired controversy, and the question of basic or derived, stable or unstable, universal or specific. It leaves room for accounting for many or few needs, long lasting needs or even changing needs. It is thought that the concept as here defined will adequately order the facts of behavior whatever they may be (6).

The seven judges were able to classify needs into 12 categories on the basis of the three criteria, the definition, and the list of needs from the literature. Of the 12 need categories, the definitions of 11 follow:

1. *Acceptance:* The need to feel that others' attitudes toward one are favorable or positive. To feel that others respect, sanction, or approve of one. To be secure in the feeling that one is a worthwhile person in another's eyes. To feel that others regard one as an equal. To feel that one is not rejected.

2. *Achievement:* The need to acquire, gain, receive, win, or strive to accomplish goals, tokens of status and respect, or knowledge. To attain, prove, secure, surmount through praiseworthy exertion.

3. *Affection:* The need to be loved, cherished, emotionally wanted for one's own sake; to receive unconditional love and affection; to receive emotional love from parents, relatives, or friends.

4. *Approval:* The need to have others' behavior toward one indicate that one is a satisfactory person, or that one's deeds are satisfactory. To seek overt rewards or other signs of approval. To be given overt demonstration by others of one's worthiness. To avoid blame, criticism, and punishment.

5. *Belonging:* The need to feel a part of a group or institution. To identify one's self with a person, group, institution, or idea. To be a member of some genial group.

6. *Conformity:* The need to be like others, to avoid marked departure from

the mode. To yield or conform to custom. To avoid being different in dress, behavior, attitudes, and ideals.

7. *Dependence:* The need to have to ask for or depend on others for emotional support, protection, care, encouragement, forgiveness, or help.

8. *Independence:* The need to be free of external control by friends, family, associates, and others. To do things in a self-determining manner, to make one's own decisions, to be self-sufficient, and to rely on one's self.

9. *Mastery-Dominance:* The need to control, to be in power, to lead, to manage, govern, control, overcome people, problems, or obstacles. To influence the behavior, feelings, or ideas of others.

10. *Recognition:* The need to be noticed, to become known. To avoid effacement of one's own individuality. To be identified by others as a unique individual, to be distinguished from others. To find one's place, to be regarded as an important human being.

11. *Self-realization:* The need to feel in sympathetic rapport with parents, relatives, friends, or associates. To feel one with others. To feel free to express one's innermost thoughts and problems to one or more persons without loss of affection or personal status. To feel that another identifies with oneself.

To obtain a representative picture for each need category, items were constructed representing each of eight behavioral situations or environmental settings. Each of the 12 need categories was presented in the following environmental context: 1) Self, 2) Unisexual peer group, 3) Heterosexual peer group, 4) Unisexual pairs, 5) Heterosexual pairs, 6) Family, 7) School, 8) Adults.

12. *To Be Understood:*

The item design was an adaptation of that used by Rogers in *A Test of Personality Adjustment* (12). Each item presented a statement about hypothetical boys or girls behaving in hypothetical settings; for example:

Item No. 1. Luke almost always stays out of trouble.
Am I like him? Yes No
Do I want to be like him? Yes No

The assumption is made that a Yes-Yes response (Yes, I am like him, and Yes, I would like to be like him) or a No-No response (No, I am not like him, and No, I do not want to be like him) indicates satisfaction with an existing state of affairs (the need is either being fulfilled, or the need is not an issue—i.e., there is present a lack of dissatisfaction). On the other hand, a Yes-No response (Yes, I am like him, and No, I don't wish to be like him) or a No-Yes response (No, I am not like him, and Yes, I would like to be like him) represents dissatisfaction with an existing state of affairs and indicates that a perceived need is not being fulfilled.

The inventory was originally administered to the entire high-school population in Delaware, Ohio. Murray (11) has compiled normative information for each item, psychological variable (need category), behavioral setting, and for the inventory as a whole. The tabulations are presented by age and sex of respondents. Weinberg (15) investigated the statistical significance for acceptance or rejection of needs by the Delaware adolescents grouped by age and sex.

2. *Thematic Picture Stories Test*

To measure differences in aggression between the delinquents and nondelinquents, a thematic picture technique was employed. Four TAT cards (3BM, 4, 6BM, and 7BM) and six original picture cards were used. The original picture cards may be described as follows:

Card A: A boy sitting in an office with his head resting on his hand (photograph from a magazine).

Card B: A blurred photograph showing two boys in movement. A rear view of a running man (from a magazine photograph).

Card C: A boy and a girl sitting on a park-bench (a magazine photograph).

Card D: A boy seated with his hands on his head. Books are piled in front of him. One of the books is opened (pencil and ink drawing by Mr. T. H. Bradrick).

Card E: Boy leaning on lamppole. Man and woman standing facing the boy. The man is pointing his finger at the boy (pencil and ink drawing by Mr. T. H. Bradrick).

Card F: Group of boys. One has his clenched fist cocked. Another boy in the foreground has his head tilted (pencil and ink drawing by Mr. T. H. Bradrick).

All 10 thematic cards were administered to individual subjects following a 15-minute interview to secure a satisfactory working relationship between the subject and experimenter.

Directions to the subject were read from the TAT manual[1] as follows:

> This is a story-telling test. I have some pictures here that I am going to show you, and for each picture I want you to make up a story. Tell what has happened before, and what is happening now. Say what the people are feeling and thinking and how it will come out. You can make up any kind of story you please. Do you understand? Well, then, here is the first picture.

---

[1] Harvard University Press, Cambridge, Massachusetts, 1943.

Verbatim typescripts were made from tape recordings of the stories produced by the *S*s. Pauses of longer than five seconds were indicated in the typed protocols. Identical face sheets attached to the typed protocols of the TAT stories and original picture stories contained the subject's code number, age, school grade, intelligence test score, race, and family status (divorced, father or mother deceased, or intact), and number of siblings. The name of the subject and his delinquency status were withheld.

Each story was scored by three judges using Jensen's system (5), a 10-point scale for rating aggressive content on the TAT. Examples of the criteria for extreme scores in Jensen's system are indicated as follows: (*a*) zero—inert, no action, mere contemplation, neutral or nonaggressive action, no indication of aggression in any form; and (*b*) 10—physical aggression resulting in death, murder, killing (killing or intent to kill must be explicit).

E. Analysis of Data

Differences in need-satisfaction patterns of delinquent and nondelinquent *S*s were determined by item analysis of responses on the needs inventory. Differences in the proportion of *S*s responding Yes-Yes, Yes-No, No-Yes, and No-No were tested for significance by Chi square. Differentiating items were then classified as to psychological variable, behavioral setting. A difference in need for goal attainment would be indicated by a difference in the proportion of Yes-No and No-Yes responses (disequilibrium) as compared to the proportion of Yes-Yes and No-No responses (equilibrium), a preponderance of Yes-No and No-Yes responses being assumed to indicate a need for goal attainment.

The degree of aggressive content for the thematic stories was rated by four judges according to the criteria established by Jensen (5). The judges were school psychologists, experienced in examining delinquent boys. Two of the judges rated half of each group.

Scorer reliability was estimated for all 10 stories—the four TAT stories and the six original picture stories—using the Coefficient of Concordance, W. Finally, differences between delinquents and nondelinquents on aggression scores were evaluated using the t test for matched pairs.

F. Results

1. *The Needs Inventory*

Items of the needs inventory discriminating between delinquent and comparison groups on the basis of differences in proportion of subjects selecting the Yes-Yes response pattern (needs are being satisfied) are shown in Table 1.

TABLE 1
ITEMS DISCRIMINATING BETWEEN DELINQUENT AND COMPARISON GROUPS BY DIFFERENCE IN FREQUENCY OF YES-YES RESPONSE

Item No.	Frequencies Del.	Comp.	Proportions Del.	Comp.	Chi square[a]	Phi coefficient
1	9	32	.33	.58	4.5	.23
6	11	10	.41	.18	4.8	—.24
11	9	36	.33	.65	7.6	.30
20	13	40	.48	.73	4.8	.24
22	13	40	.48	.73	4.8	.24
24	9	35	.33	.64	6.7	.29
30	7	26	.26	.47	3.4	.20
38	7	27	.26	.49	4.0	.22
39	7	31	.26	.56	6.5	.29
43	12	38	.44	.69	4.6	.23
46	13	39	.48	.71	4.0	.22
50	11	33	.41	.60	4.5	.23
56	10	36	.37	.65	5.9	.27
57	7	31	.26	.56	6.5	.28
59	6	23	.22	.42	3.7	.21
62	13	41	.48	.75	5.6	.26
70	11	39	.41	.71	6.9	.29
73	6	39	.22	.71	17.3	.46
75	10	36	.37	.65	5.9	.27
76	9	37	.33	.60	8.5	.32
77	10	35	.37	.64	5.1	.25
78	8	36	.30	.65	9.3	.34
82	7	30	.26	.55	6.0	.27

[a] The distribution for Chi square ($df = 1$) follows: $p < .10$, $X^2 = 2.7$; $p < .05$, $X^2 = 3.8$; $p < .02$, $X^2 = 5.4$; $p < .01$, $X^2 = 6.6$; $p < .001$, $X^2 = 10.8$.

There are 23 discriminating items. On 22 of these items a smaller proportion of delinquents than nondelinquents select the Yes-Yes pattern, indicating that delinquents display a distinct tendency to be dissatisfied with the way things are—or at least to wish they were different—and that delinquents may be assumed to display a greater need than nondelinquents for goal attainment. All items shown in Table 1 with the exception of items number 30 and 59 are significant beyond the .05 level. Items 39 and 59 are significant at the .10 level.

The fourfold point correlation for the 23 items ranges from 120 to 146 with a median coefficient of .26. Interpretation of the Phi coefficient is difficult because the values for Phi are considerably lower than the tetrachoric correlation coefficient. The obtained values indicate a low to moderate relationship between delinquency status and the proportion of subjects selecting the Yes-Yes response pattern.

Table 2 indicates items discriminating between delinquent and comparison groups on the proportion selecting the No-No, Yes-No and No-Yes response

TABLE 2
ITEMS DISCRIMINATING BETWEEN DELINQUENT AND COMPARISON GROUPS ON NO-NO, YES-NO, AND NO-YES RESPONSE PATTERNS

Response pattern	Item No.	Frequencies		Proportions		Chi[a] square	Phi coefficient
		Del.	Comp.	Del.	Comp.		
No-No	18	13	38	.48	.69	3.3	.21
	22	10	8	.37	.15	5.5	—.26
	30	9	8	.33	.15	4.0	—.22
	31	4	19	.14	.35	3.5	.21
	41	5	22	.19	.40	3.8	.21
	44	2	20	.07	.36	7.3	.30
	74	8	38	.30	.69	11.5	.37
Yes-No	38	11	2	.41	.04	4.0	—.22
	53	6	25	.22	.46	4.2	.23
No-Yes	10	3	22	.10	.40	7.4	.30
	44	12	11	.44	.20	4.8	—.24
	73	13	10	.48	.18	7.9	—.31
	83	7	27	.27	.49	4.0	.22

[a] The distribution for Chi square ($df = 1$) follows: $p < .10$, $X^2 = 2.7$; $p < .05$, $X^2 = 3.8$; $p < .02$, $X^2 = 5.4$; $p < .01$, $X^2 = 6.6$; $p < .001$, $X^2 = 10.8$.

patterns. As in the case of the Yes-Yes pattern, the Phi coefficient indicates a low to moderate relationship between response pattern and the delinquency status. Many items had to be excluded from the analysis on the No-No, Yes-No and No-Yes patterns because the total number of *S*s responding was small, leading to an expected frequency of less than five in the Chi-square analysis. Since there is some disagreement among statisticians as to the justification for using very small numbers, it was decided to adopt a conservative strategy and exclude items of this description from analysis.

The analysis of response patterns indicates that the most striking differences between the delinquent and the nondelinquent groups is on the Yes-Yes pattern. With only one exception in the 23 discriminating items, fewer delinquent than nondelinquent subjects select this response mode. The nondelinquent subjects indicate that the hypothetical roles presented in the needs inventory are descriptive of their behavior and that they are satisfied with those roles. On the other hand, delinquent *S*s indicate that the hypothetical roles are inappropriate descriptions of their behavior or that they do not accept the hypothetical roles.

Twelve need categories were devised by Lucas and his colleagues as a way of classifying the directions of adolescent motivations. In addition to the "needs" or "psychological variables," Lucas considered the situational aspects of behavior to be an important variable. Thus, items for the revised needs questionnaire were constructed on the basis of *a priori* variables and situations.

In Table 3, items found to discriminate between the delinquent and the comparison group on the basis of one or more of the response patterns are classified by the behavioral situation and psychological variable. Table 3 should be consulted in relation to Tables 1 and 2, the item numbers shown in all three tables being the same, with Tables 1 and 2 indicating the discriminative value of the item and the direction of its discrimination.

The majority of the items shown in Table 3 show nondelinquents as expressing satisfaction with their need status, and delinquents as showing significantly less satisfaction (i.e., greater need for goal attainment). Seven of the eight items included in the psychological variable "approval" significantly discriminated between delinquent and nondelinquent groups, indicating delinquents as being more dissatisfied with their attainment of their "approval" need than were nondelinquents. Four of the eight items included in the variable "self-realization" also discriminated in the same direction. The remaining 21 items are spread among the other 10 psychological variables. Examination of the behavioral situations indicates the "family" setting to be most important in differentiating between delinquent and nondelinquent responses. Eight of the 12 items relating to the "family" showed significant

TABLE 3
ITEMS DISCRIMINATING BETWEEN DELINQUENT AND COMPARISON GROUPS
CLASSIFIED BY BEHAVIORAL SITUATION AND PSYCHOLOGICAL VARIABLE

Variable	Behavioral situations[a]							
	1	2	3	4	5	6	7	8
				Items				
Accept								50
Achievement		62	83					
Affection								76
Approval	1	20	30	41		78	82	53
Belonging		10				70	73	
Conformity						57	74	
Dependence		59				44		
Independence		11	18			75		
Mastery-Dominance		46				6		34
Recognition							31	39
Self-Realization	56		77			43		22
To Be Understood						24		38

[a] Key to Behavioral Situations:
1. Self
2. Unisexual peer group
3. Heterosexual peer group
4. Unisexual pairs
5. Heterosexual pairs
6. The family group
7. The school
8. Adults.

differences in the response patterns of the two groups with only one (Item 6) of the eight items indicating less need for goal attainment on the part of the delinquents. Seven of the 12 items included in the "adult" classification and half of the 12 items relating to "unisexual peer groups" were found to be significant. Table 4 presents the content of the items discriminating between delinquent and nondelinquent groups.

2. *Thematic Picture Stories Test*

Scorer reliability was estimated using Kendall's Coefficient of Concordance, reliability being approximately the same for the TAT and picture stories

TABLE 4
CONTENT OF ITEMS DISCRIMINATING BETWEEN DELINQUENT AND NONDELINQUENT (COMPARISON) GROUPS

Item No.	Item
1	Luke almost always stays out of trouble.
2	Charles can often make his parents change their minds.
10	Bill is always welcome to go around with a bunch of boys he knows.
11	Mac doesn't always do what the other boys in his gang do.
18	Ted goes ahead and does some things that the other boys and girls think he shouldn't do.
20	Other boys often ask Red to do things with them.
22	James doesn't want a job where he will loaf.
24	Harvey can talk to his family about his thoughts and problems.
30	The boys and girls seldom say anything bad about Albert.
31	Jim is president of quite a few clubs.
38	Older people usually know how Andrew feels about things.
39	Everybody says Bob will do very well for himself after he finishes school.
41	Joe has a boy friend who doesn't like some of the things Joe does.
43	Jerry has a chance to show how well he really can do.
44	Louis goes to his family for advice; he doesn't decide many things all by himself.
46	The other boys listen to Hal when he makes suggestions.
50	Older people like to have Sidney around.
53	Older people sometimes treat Ralph as if he were a younger boy.
56	Nothing stops Hugo from doing the best he can.
57	Walter does what his parents expect him to do.
59	Lee gets help, sympathy, and advice from the boys in his crowd.
62	Herb usually succeeds when he works hard for what he wants.
70	Pete and his family enjoy doing many things together.
73	Bert goes to one of the best schools in this part of the state.
74	Fred sometimes breaks school rules because the other boys do.
75	Vic's family let him make up his own mind about many things.
76	Older people think a lot of Howard.
77	Instead of just getting by, Adam tries to do his best.
78	Paul's family approves of most things he does.
82	Arthur's teachers give him praise now and then.
83	Craig can date almost any of the girls he knows whenever he tries.

Note: The needs inventory consisted of 90 items. Of these the above 31 discriminated significantly between delinquent and nondelinquent groups in terms of need for goal attainment. The remaining 59 did not discriminate at any acceptable level of confidence.

(*W* equals .55). Scorer agreement was also similar for delinquent and nondelinquent stories.

The mean aggression scores for the delinquent group were lower than for those of the controls, on the TAT Pictures, picture stories, and complete 10-picture instrument, as shown in Table 5—indicating a trend for the delinquents to express less aggression in their stories. To take advantage of the pairing of subjects, the t test for matched pairs was used to evaluate the significance of differences in aggression scores. Table 6 indicates that the 10-picture instrument discriminated between groups. Differences in aggression scores were lower for the delinquent subjects on both the TAT and picture stories. However, the results of the significance tests indicate that the differences in ratings were too small to be significant for the limited samples of responses obtained on the six picture stories and the four TAT pictures considered separately. On the larger sample obtained for the 10-picture instrument, the difference in aggression scores between delinquents and nondelinquents was significant at the .05 level. The biserial correlation coefficients shown in Table 7 also support this conclusion. These data indicate that the

TABLE 5
DESCRIPTIVE STATISTICS FOR DELINQUENT AND CONTROL GROUPS

Instrument	Mean aggression score Del.	Mean aggression score Control	Standard deviation Del.	Standard deviation Control
4 TAT pictures	45.6	52.2	17.0	18.4
6 Picture stories	78.0	85.1	18.4	15.8
10 Pictures	123.1	137.3	26.9	28.1

TABLE 6
DIFFERENCES BETWEEN DELINQUENT-NONDELINQUENT MATCHED PAIRS ON AGGRESSION SCORES

Instrument	Sum D	Sum D^2	t
4 TAT pictures	—152	7673	1.83
6 Original pictures	—139	10307	1.41
10 Pictures	—291	19899	2.26*

* Significant at .05, two-tailed test.

TABLE 7
BISERIAL CORRELATIONS BETWEEN TOTAL AGGRESSION SCORES AND DELINQUENCY STATUS

Instrument	Biserial r
4 TAT pictures	—.24
6 Picture stories	—.26
10 Pictures	—.32

relationship between delinquency status and aggression scores was consistently negative. The correlation is low although not negligible.

G. Discussion

Of major theoretical interest is the finding that the items discriminating between the delinquent and the nondelinquent subjects on the revised needs questionnaire were grouped more decisively by the "behavioral settings" classification than by the "psychological variable" or "need category" classification. It is evident that there is a piling up of items on three of the eight "behavioral settings," while only one "need category" is clear in accounting for the items. This finding, should it be other than an accidental occurrence, would imply that the place to start looking for differential behavior of delinquents should be in settings in which the behaviors take place. One setting, the home of the delinquent, stands out as a key to an understanding of delinquent behavior, according to the responses on the revised needs inventory confirmed by the extensive literature on delinquency. Unfortunately, much of the information about the family life of the delinquent youth reported in the literature is from interviews with the delinquent and his parents. The accuracy of this information and interpretations based on it are open to serious question. The present findings suggest that something is awry in the family relationships of the delinquent. While both interview and inventory evidence give indirect support to disturbed family situations as the focus of delinquent behavior, the need for direct observations of the family life of the delinquent is evident.

The finding of a negative correlation between delinquency and aggression ratings on the thematic picture stories supports hypotheses made by Symonds (14) and by Sanford (13) that there is an inverse relationship between needs expressed on the TAT and "overt" behavior. If this finding is confirmed in future studies, it would suggest that the TAT might be used in the measurement and prediction of behavior, a consummation desired by adherents to and critics of picture projective techniques.

The approach in this study, as in most studies of delinquency, was to find differences between delinquents and nondelinquents. While differences were found, the similarity of the groups should not be overlooked. While one-third of the items on the revised needs questionnaire dscriminated between delinquents and nondelinquents, two-thirds did not. The discriminating items clearly "piled-up" in only one "need category," suggesting that motive organization is very similar for both groups. While significant differences in aggressive content in stories were found, they were small. The evidence suggests that the classification "delinquent" has more to do with the environmental setting than with individual motive patterns measured apart from that setting.

H. Summary

From Miller's (2) statement of the frustration-aggression hypothesis, two hypotheses were proposed as a means of studying differences between delinquent and nondelinquent adolescent males. It was hypothesized that delinquents and nondelinquents differ in degree of goal attainment and that they differ in "aggression."

Differences in goal attainment were studied in an item analysis of responses on an adolescent needs questionnaire. The Chi square test at the .05 level was used to evaluate differences in the proportions of delinquent and nondelinquent subjects selecting alternative response categories. Discriminating items were classified according to Lucas' (6) system of psychological variables and behavioral settings.

Differences in aggression were investigated using a 10-picture thematic instrument composed of four TAT pictures and six original pictures. Each story was rated independently by four judges who were not informed of the delinquency status of the subjects. In addition, intensity of aggressive content was rated on a six-point scale devised by Jensen (5). Differences in total aggression scores for delinquent-nondelinquent pairs were evaluated by the *t* test for matched pairs. The .05 level of significance was adopted.

The needs questionnaire was administered to 27 delinquent males held in a county detention home and to 55 nondelinquent high-school males. The groups were similar in age and intelligence.

The original subjects were not available for use with the thematic materials, necessitating the drawing of a second sample for this instrument. Thematic pictures were administered to 21 pairs of delinquent and nondelinquent males, individually matched on age, intelligence and race. All subjects resided in the same high-school district, partially controlling for socioeconomic status. Differences between delinquents and nondelinquents were found for 32 items on the 90-item needs questionnaire, indicating a clear tendency for delinquents to be less satisfied with their status (expressed in terms of needs) and hence to display a greater need for goal attainment than was true of nondelinquents, who were more satisfied with their status. Seven of the eight items comprising the psychological need "approval" discriminated between delinquents and nondelinquents. Evidence of thwarting in goal attainment was found in behavioral settings involving the family group, adults, and the unisexual peer group.

On the thematic-picture material, delinquents and nondelinquents differed in intensity of aggression, with delinquents receiving lower ratings than nondelinquents on the four TAT stories and the six original stories, and hence showing a tendency to display less aggression than did nondelinquents in their stories. For the 10-picture instrument, the mean of differences between matched pairs was significant at the five per cent level.

The hypothesis that delinquents and nondelinquents differ in goal attainment was supported by the findings of this study. The hypothesis that delinquents and nondelinquents differ in "aggression" as defined in the investigation was also confirmed.

REFERENCES

1. BENIMOFF, M., & HORROCKS, J. E. A developmental study of relative needs satisfactions of youth: Equilibrium—disequilibrium. *J. Genet. Psychol.*, 1961, **99**, 185-207.
2. DOLLARD, J., MILLER, N. E., DOOB, L. W., MOWRER, O. H., & SEARS, R. R. Frustration and Aggression. New Haven, Conn.: Yale Univ. Press, 1939.
3. HEALY, W., & BRONNER, A. F. New Light in Delinquency and Its Treatment. New Haven, Conn.: Yale Univ. Press, 1936.
4. HORROCKS, J. E. The Psychology of Adolescence (2nd ed.). Boston: Houghton Mifflin, 1962.
5. JENSEN, A. R. Aggression in fantasy and overt behavior. *Psychol. Monog.*, 1957, **71**, No. 16 (Whole No. 445).
6. LUCAS, C. M. An emergent category approach to the analysis of adolescent needs. Ph.D. dissertation, Ohio State University, Columbus, 1951.
7. ———. Supplementary research for an emergent category approach to the analysis of adolescent needs. Mimeographed paper, Cedar Crest College, Allentown, Pennsylvania, 1955.
8. LUCAS, C. M., & HORROCKS, J. E. An experimental approach to the analysis of adolescent needs. *Child Devel.*, 1960, **31**, 479-487.
9. LUCAS, C. M., HORROCKS, J. E., *et al.* Youth needs questionnaire. Mimeographed paper, Ohio Youth Survey Associates, Ohio State University, Department of Psychology, Columbus, 1951.
10. MERRILL, M. A. Problems of Child Delinquency. Boston: Houghton Mifflin, 1947.
11. MURRAY, R. L. Adolescent needs: A cross-sectional community research. Master's thesis, Ohio State University, Columbus, 1952.
12. ROGERS, C. R. A Test of Personality Adjustment. New York: Association Press, 1931.
13. SANFORD, R. N., ADKINS, M. M., MILLER, R. B., COBB, E. A., & OTHERS. Physiques, personality and scholarship: A cooperative study of children. *Monog. Soc. Res. Child Devel.*, 1943, **8**, No. 1 (Serial No. 34).
14. SYMONDS, P. M. Adolescent Fantasy: An Investigation of the Picture-Study Method of Personality Study. New York: Teachers College, Columbia Univ. Bur. Publ., 1949.

15. Weinberg, S. A. Psychological needs of adolescence. Ph.D. dissertation, Ohio State University, Columbus, 1953.

Department of Psychology
The Ohio State University
Columbus, Ohio

Department of Psychology
University of Minnesota
Minneapolis, Minnesota

Published as a separate and in *The Journal of Psychology,* 1966, **62**, 195-200.

POTENTIAL VALUE OF EDUCATIONAL BACKGROUND DATA IN THE SELECTION AND CLASSIFICATION OF MILITARY PERSONNEL*

Personnel Research Laboratory, Lackland Air Force Base

CHESTER J. JUDY

A. INTRODUCTION

Service and business authorities have generally been sold testing programs on the basis of "improvement over chance" computations, in which chance selection from an applicant population is assumed. But as Cronbach and Gleser (1) have pointed out, chance selection is not the only alternative to a selection program based upon the use of tests. Well-constructed examinations will sometimes improve decisions, but not as much as is implied by the phrase "improvement over chance." If testing were impossible in military or industrial selection, it would not be necessary to fall back upon chance decisions because other data could be used with substantial predictive validity. Just how substantial some of these values are, or can be, has not been documented. The cost or efficiency of few testing programs has been evaluated in terms of what can otherwise be accomplished with modern data-handling equipment, using "free" or already-available information. Particular attention is here being given to a limited number of educational background variables and their potential as predictors of training criteria.

Although the orientation of the material in this report is toward the resolution of an issue having to do with selection and classification strategies in military and industrial settings, it also bears upon a current academic matter concerning the utility of general education as background for vocational training or placement.

B. PROCEDURE

During a search to identify promising variables for the purposes of this investigation, a choice was made to examine the kind of information that is routinely assembled on service people, or that information which can easily be obtained. For incoming officer personnel for the Air Force, major subject

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 14, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

in college, overall grade point average (GPA), and name of college attended are among the several educational background data that are made a matter of record. In the analyses reported below, eight variables in various combinations from four categories of such information are related to technical-school performance. Some of the data analyzed was obtained through the use of a short questionnaire. Eight educational variables were used in each analysis because eight subtests from the Air Force Officer Qualifying Test (AFOQT) are used in deriving the Navigator-Technical Aptitude Composite (NTAC). Full advantage was taken of electronic data processing procedures in generating relevant educational variables. It was, for example, thought important to study what a student officer had not majored in, in college, as well as what he *had* majored in. While a "College Code" was being developed, consideration was given to the circumstance that there is probably some variation in the caliber of student officers coming to the Air Force from different colleges, that this information is contained in Air Force records, and that the information can be used to help predict success in technical training. The use of a College Code with other educational data (major, GPA) was prompted, partly, by a remark of Cronbach and Gleser (1, p. 125) in which Professor P. E. Vernon is quoted as saying that in England success in grammar school has been predicted from properly scaled primary-school ratings with validity coefficients as high as .80. The GPA of an officer from College A may, in fact, mean considerably more (or less) than the same average of another officer from College B. It probably makes some difference, too, just where a given major field of study was pursued, if major subject information is to be used in a prediction equation.

The data used in the analyses reported here were a part of the material assembled in connection with a larger and continuing study designed to identify measures which serve to differentiate student officers in terms of their successfully completing various Air Force technical training courses. Individuals completing basic technical courses between April 1964 and April 1965 were asked to give estimates concerning the number of semester hours of college credit earned in each of 14 subjects (physics, chemistry, meteorology, other physical science, mathematics, aeronautical engineering, electrical engineering, mechanical engineering, chemical engineering, civil engineering, other engineering, military science, business administration or management, and social science). These estimates were taken as auxiliary data and used with information already on hand in Air Force personnel files to predict final school grade (FSG) in Air Force courses. Correlational techniques were used in establishing a series of relationships between the FSG criterion and eight variables generated from increasing amounts of educational information. Results are given for five Air Force courses, those for which the largest *N*s can be reported.

C. RESULTS

Results from the correlational analysis of five sets of data are presented in Table 1. Listed on the left are the title and number of five Air Training Command Courses. Following the course number, within parentheses in the same column, is shown the number (*N*) of student officers on whom data were obtained.

The second column shows the relationship between the final school grade in the technical course and college major information. The multiple correlation values shown here were arrived at after an arbitrary limit of 26 possible major fields of study was established. (Combinations such as Psychology/Sociology or Philosophy/Religion were arbitrarily made to reduce the number of necessary categories.) Twenty-six variables were generated to allow the individual student officer to be identified in terms of his major subject in college. (A "major" variable was assigned the value of "1" if it represented an officer's college major, or "0" if it did not.) Eight of these variables—those which, in combination, showed the highest relationship with the FSG criterion—were taken as "good" major variables. Table 1 values in the second column are multiple correlations between these eight variables and FSG. The third column in the table shows the multiple correlations after GPA in all college work, as a variable, had been given a chance to compete with all "major" variables as a good predictor of the FSG criterion. Again, the number of educational variables related to the FSG criterion, in arriving at the values shown in the table, was limited to a total of eight.

The values shown in the fourth column of Table 1 (College Code information added) were arrived at after two additional educational variables were generated and given a chance to compete with those already described. These additional variables were arrived at by assigning, to each college represented, a value equal to (*a*) the mean score of all other student officers from the given college on the eight AFOQT subtests from which the NTAC is computed, and (*b*) the mean score of all other officers on three subtests from which the Quantitative Aptitude Composite is computed. The fourth column contains multiple correlation values obtained when information on number of semester hours of college credit earned in each of 14 subjects was first used to generate a set of variables which were allowed to compete with the previously described variables generated from the available educational data. The number of educational variables related to the FSG criterion in arriving at Table 1 values was in each instance limited to a total of eight.

The multiple correlation values shown in the last column of Table 1 indicate the relationship between the FSG criterion and the eight AFOQT subtests from which the NTAC is computed. These last values are a few correlation points above the values obtained when the eight subtests are combined into a single NTAC score, the score actually used in Air Force selection and classification operations. (With present data, the range is from .411 to .649 rather than from .396 to .621.)

TABLE 1
RELATIONSHIP† BETWEEN FINAL TECHNICAL-SCHOOL GRADE AND (*a*) SELECTED INFORMATION ON COLLEGE EDUCATION AND (*b*) CLASSIFICATION TEST DATA‡

Air Training Command Course	College major	Major and GPA	Major, GPA, & College Code	Plus sem. hrs., 14 subjects	Test data
Aircraft Maint. Officer, OBR4341 ($N = 450$)	.302	.360	.411	.460	.448
Communications Officer, OBR3031 ($N = 260$)	.444	.489	.504	.547	.649
Munitions Officer, OBR3271 ($N = 126$)	.479	.657	.674	.737	.543
Weapons Controller, OBR1741 ($N = 131$)	.390	.422	.474	.537	.337
Supply Officer, OBR6421 ($N = 191$)	.320	.419	.466	.472	.411

† Values in table are multiple correlation coefficients showing the relationship between final school grade and eight variables generated from information of the kind indicated in column headings.

‡ Eight subtests from the Air Force Officer Qualifying Test (AFOQT). These subtests are the ones used in deriving the "Navigator-Technical Aptitude Composite," a score used in the selection and classification of officer personnel for technical training in important specialties.

Table 1 values for the available college education information (next to last column) are, for all courses except one, larger than those shown for the test variables (last column). For three of the five courses, information on college major and overall GPA turned out to have predictive validity equal to or greater than that reported for the test data. In each instance, improvements over the validity reported for college major information are realized when GPA is taken into consideration. A further increase, although not always large, is realized when College Code information is taken into account. A further increase in the validities reported is seen when questionnaire data are used.

All the values in Table 1 are large enough to be statistically significant at the .01 level. A few of the increases shown in adjoining columns for the educational data are not large, but every table entry except one (pertaining to the educational data) that shows an increase over the one in the previous column reports a value significantly larger (.05 level of statistical significance) than one computed with the newly added variables deleted. In the Supply Officer group, the increase in validity is from .455 to .472, not enough to be statistically significant. (The value .455 should not be confused with the table value of .466. The former is the multiple correlation between FSG and seven educational variables, not eight.)

D. Discussion

In view of the circumstance that a rather incidental amount of educational data (most of it already on hand) has been examined, it is probably true that the practical limits to the predictive validity of this kind of information has not been reached. It is inconceivable, for example, that grade point average in particular college subject matter areas would not be more useful than overall GPA of the gross kind used in the present analyses. Very little effort has, in fact, been devoted to the study of background educational variables, as contrasted with the amount of effort devoted to the study and refinement of variables obtainable from the administration of written tests. The latter has extended over a period of several decades in the Military Services, leading McNemar (2, p. 873) to observe what he interprets as a hope "that some jobs will have aptitudes in common so that the needed number of tests will be less than the number of jobs." Travers (3, p. 198), noting the effort devoted to the building and validating of new tests, suggested that "to the test technologist the situation may appear to be one of despair for the evidence would indicate that the limit of progress has been reached."

Some of the difference in favor of educational information observed in Table 1 may be explained on the premise that more attention was given to test scores than to background educational information when selection and assignment decisions (with respect to the present officer groups) were made. An examination of the basic data does suggest that individuals comprising these groups were not particularly well qualified, educationally. Less than 15 per cent of the 450 student officers preparing for the Aircraft Maintenance Officer specialty, for example, had the "desirable" educational qualifications, as these qualifications are given in Air Force documents. The use of test scores, to the exclusion of other important information, may in many instances operate to prejudice good assignment decisions.

Another uncertainty with respect to the relative magnitude of the validities reported here is that the eight educational variables used in each analysis were chosen from a number larger than eight. This procedure capitalizes upon chance fluctuations among the relationships studied. There was, however, some regularity in the manner in which three or four variables served as good predictors across all groups. Before any final degree of importance can be claimed for the remaining variables, other studies will have to be made using larger *N*s than those for any of these groups except the Aircraft Maintenance Officer group. It does seem likely that educational variables may have some differential utility as predictors of success in different specialties.

There are plausible explanations why educational variables of the kind used in the analyses reported here can be expected to approach or surpass the best combination of relevant AFOQT subtests as useful predictors of initial success in the Air Force. Partly, perhaps, it is because the college record represents the cumulative observation of a number of moderately bright persons over a substantial period of time, and because test-like situations prevail as a matter of custom. To some extent, college records may reveal abilities, interest patterns, and motivations of the important variety not yet successfully measured for the purposes of military selection and classification.

E. SUMMARY

Correlational techniques were used to examine the relationships between selected educational background variables and final school grade in five technical courses conducted for Air Force student officers. The conclusion was reached that the relationships are high enough to demonstrate the general usefulness of background information on college education for predicting technical school performance, and that the validity of educational information may be expected to approach or surpass that which can be demonstrated for a current selection and classification test.

REFERENCES

1. CRONBACH, L. J., & GLESER, G. C. Psychological Tests and Personnel Decisions. Urbana, Ill.: Univ. Illinois Press, 1957.
2. MCNEMAR, Q. Lost, our intelligence? Why? *Amer. Psychologist,* 1964, **19**, 871-882.
3. TRAVERS, R. M. W. Personnel selection and classification research as a laboratory science. *Educ. & Psychol. Meas.,* 1956, **16**, 195-208.

106 Thomas Jefferson Drive
San Antonio, Texas 78228

Published as a separate and in *The Journal of Psychology*, 1966, 62, 201-203.

A NOTE ABOUT COMMUNICATION RESEARCH[*]

Research Department, Campbell-Ewald Company

DANIEL K. STEWART[1]

Any discipline that purports to be of a scientific nature seeks to establish as a result of its research an organized body of theory, law, and description. Science is the pursuit of truth for its own sake, and this product, where approximated, is always explicated in the form of generalized laws or descriptive propositions. These propositions form coherent systems of knowledge, and such systems provide for the explanations and understanding of the phenomena of which they deal. Nothing else is pertinent, because "nothing else" is testable.

The subject matter of communication research is clearly in need of some type of organization along these lines. We are in immediate need of both a philosophy adequate in providing a theoretical foundation valid for matters of the human mind, and a psychology commensurate with such a philosophy. If the explanatory power of a scientific discipline is measured in terms of its theories, laws, and descriptions as is indicated by Feigl (1) and by Hempel and Oppenheim (3), then communication research must bring its data and statistical tabulations to bear on the very products of the communication process—namely, given ideas and given meanings.

* * *

One scientific law of human communication is this: communication entails the existence of some physical symbol complex (usually words), its intended meaning, and some mind interpreting that complex. This law implies the specification that communication, in fact, exhibits physical, engineering, or vehicular factors. Since "meaning" is a function of ideas, this law implies the specification that communication, in fact, employs these physical vehicles for the very purpose of eliciting *ideas* in the minds of interpreters. Finally, this law implies that significant communication research must account for these ideas in its experimental designs.

---

[*] Received in the Editorial Office, Provincetown, Massachusetts, on February 14, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Dr. Stewart received his degree from Michigan State University in 1959, and is a former member of its faculty. Now with the Campbell-Ewald Company in Detroit, he is responsible for research in the sciences and arts pertaining to communication.

However, there is much in communication research today which seems to be directed away from these products of the communication process. One such enterprise goes under the title of Information Theory, but Information Theory is relevant to the communication of ideas only if it can bring its data to bear on that relation between physical complexes (e.g., words) and ideas. We have to resolve the psychological question of when do physical symbol complexes become "information"?

We cannot assume, *a priori,* that "information" and "sentences" are synonyms. To say that " 'information' . . . must not be confused at all with meaning . . . information is a measure of your freedom of choice when you select a message" (7, p. 12) is to prejudge the experimental results when the term "message" entails "meaning" in its scope of intention.

From a purely psychological point of view, the measurement of "choices" would seem to necessitate an analysis of those meanings which different symbol complexes elicit. Only in this way could "Symbol Complex A" be differentiated experimentally as a choice from "Symbol Complex B."

It is said that "this term (information) . . . is not synonymous with 'meaning.' Only the amount of information is measured—the amount does not specify the content, value, truthfulness, exclusiveness, history, or purpose of the information" (4, p. 3). We note that this remark has not yet explicated what "information" means.

Specifically, we have to know the nature of this "information," the amount of which is being measured. Experimentally, we have to know whether this "information" consists of ideas or is purely physical. If it is ideational, then it becomes a candidate for meaning (5). To have information is to possess it mentally, as ideas; but, if the measurement of "meaning" is excluded from our experimental designs, then the objects of information must literally be something of a nonpsychological nature; because if information does not exist in the mind, where does it exist? To be quite precise, if "information" is specifically defined as excluding meaning, does this not necessarily imply that information about Information Theory is meaningless information?

* * *

Finally, logical models of communication which fail in explicating the relation between the object language of some logistic system *and* meaning would provide similar psychological difficulties (2). The nominal definitions and "well-formed formulas" of such a system stand, in fact, as metalanguage *about* communication rather than as an object language *of* communication. Therefore, the implication that such models are epistemologically relevant to the nature of communication *per se* appears to be an experimentally untenable position. In such a case, it does not follow that the meaning of "meaning" is entailed either in the set of primitive propositions or is logically implied by such a set; but it is the elicited "meaning" resulting from a communicative act which is the object of measurement. As pertinent as validity patterns (or well-formed formulas) are to the communication process, to identify them as "meaning" is to force an equivocation of "well-formed formulas" with "ideas" (6).

* * *

With respect to scientific methodology, one of the laws which governs communication research is the one stated at the beginning of this paper—namely, that communication entails the existence of some physical symbol complex (usually words), its intended meaning, and some mind interpreting that complex. This law implies that communication research is concerned with those relations involving *symbols, meaning,* and *mind* as integral factors. Any psychological explanation of communication lies in the logical conjunction of true propositions explicating such relations.

REFERENCES

1. FEIGL, H. Some remarks on the meaning of scientific explanation. In H. Feigl & W. Sellars (Eds.), *Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 510-514.
2. HARRAH, D. Communication: A Logical Model. Cambridge, Mass.: M. I. T. Press, 1963.
3. HEMPEL, C. G., & P. OPPENHEIM. The logic of explanation. In H. Feigl & M. Brodbeck (Eds.), *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts, 1953. Pp. 319-352.
4. MILLER, G. A. What is information theory? *Amer. Psychologist,* 1953, **8**, 3-11.
5. STEWART, D. K. Communication, ideas, and meaning. *Psychol. Rep.,* 1965, **16**, 885-892.
6. ———. Communication and logic: III. Evidence for showing that communication is a function of validity patterns. *J. Gen. Psychol.,* 1965, **73**, 67-72.
7. WEAVER, W. The mathematics of communication. *Sci. Amer.,* 1949, **181**, 11-15.

Research Department
Campbell-Ewald Company
General Motors Building
Detroit, Michigan 48202

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 205-209.

RAVEN'S PROGRESSIVE MATRICES: CONSTRUCT VALIDITY*

VA Hospital, East Orange, New Jersey

William C. Bingham,[1] Henry R. Burke, and Stewart Murray

A. Introduction

In a personal communication to one of the present investigators, Raven stated that the Progressive Matrices (PM) "assesses a person's present capacity for coherent perception and orderly judgment."[2] And in his *Guide to Using Progressive Matrices* (5), he states: "A person's total score provides an index of his intellectual capacity, whatever his nationality or education" (5, p. 1). Even though Raven has stated, "*By itself* it (PM) is not a test of 'general intelligence' . . ." (5, p. 2), we thought it would be useful to examine again the relationship of his PM to some other measures of general mental ability. A 1958 review and critical evaluation of the literature on PM justified the observation that inconsistencies in the findings of investigators make it extremely difficult to assert with confidence just what it is that PM measures and how it relates to age (1, pp. 202 *ff*.). Raven himself has stated that scores on PM "decline slowly (with age), but with remarkable uniformity" (5, p. 4).

Validity coefficients reported for PM ranged from .86 to .23 (1, pp. 202-207). Obtained coefficients of reliability ranged from .93 to .64 (1, pp. 214-217).

We decided to use the Army General Classification Test (AGCT), First Civilian Edition, because as "one of the most often used psychological tests" (6, p. 5), ". . . it has been shown to be an effective test of general mental ability" (6, p. 30). We also used the WAIS because Cohen, in his factor analysis of the Wechsler Adult Intelligence Scale (WAIS), identified a strong second order factor which he interpreted as "present general intellectual functioning" (3, p. 451). Hall (4) reported considerable overlap in what the PM and the WAIS subtests measure.

In addition to a strong general second order factor, Cohen (2, 3) clearly identified three major common factors: Factor A, Verbal Comprehension;

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 15, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Now at Rutgers—The State University.

[2] Raven, J. C. Letter to Henry R. Burke, Feb. 9, 1959.

Factor B, Perceptual Organization; and Factor C, Memory. In view of Cohen's description of Factor B (2, p. 285), Raven's statements about PM, and Hall's data, PM might be expected to be more highly related to Factor B than to the other Cohen factors.

B. Hypotheses

The present study set out to investigate the following hypotheses: I. PM is a measure of general intellectual functioning. II. The correlation between PM and Cohen's WAIS Factor B will be significantly greater than that between PM and Cohen's other factors. III. Performance on PM will be inversely related to age.

C. Procedure

All patients referred to Vocational Counseling, Psychology Service, at the East Orange VA Hospital, over a six-week period, to whom the relevant tests could be routinely administered, were tested by the investigators. *No time limit was used with PM.*

The *S*s ranged in age from 20 to 52 (*M,* 36.1; *SD,* 7.7). Total *N* was 39: seven from Neurology, 10 from Psychiatry, six from Medicine, five from Surgery, nine from Pulmonary Diseases, and two from Physical Medicine and Rehabilitation.

Verbal, Performance, and Full Scale Scaled Scores, as well as the three Cohen factor scores, were computed for the WAIS. Separate Vocabulary, Arithmetic, Block Counting, and Total Scores were found for the AGCT. Pearson product-moment intercorrelations were computed by electronic equipment among the 11 sets of test scores and age. They are given in Table 1.

D. Findings and Discussion

It was quite surprising with such a small ($N = 39$) and not notably heterogeneous population that no coefficient of correlation, except for age, was below .53 (Table 1). This finding tended to support the proposition that the three instruments do in fact measure much in common.

No significant differences were found between the Full Scale SS Means and Standard Deviations for our experimental subjects and Wechsler's normative age groups ($N = 300$); Mean Full Scale SS for the experimental group (age mean 36.1, *SD* 7.7) was 108.64, *SD* 23.03; for the Wechsler norm groups (ages 25-34) Mean FS SS: 110.36, *SD* 24.08; (ages 35-44) Mean FS SS: 106.3, *SD* 24.8 (7, p. 19).

With respect to our hypotheses, inspection of Table 1 indicates that the

TABLE 1
Intercorrelations of Test Score Variables and Age ($N = 39$) (Decimals Omitted)

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. WAIS Verbal SS												
2. WAIS Perf. SS	77											
3. WAIS FS SS	96	92										
4. Cohen Verbal	96	71	91									
5. Cohen Memory	76	64	75	57								
6. Cohen P. O.	67	90	82	60	58							
7. AGCT V	83	58	77	83	54	53						
8. AGCT N	89	65	84	81	76	59	82					
9. AGCT B	76	62	75	69	61	55	67	84				
10. AGCT T	89	65	84	83	70	59	90	94	89			
11. Raven	80	79	85	79	63	77	65	71	59	70		
12. Age	—14	—33	—23	—09	—19	—27	10	—12	—11	01	—12	
M	62.2	46.4	108.6	10.5	10.5	10.0	25.6	27.5	25.2	102.2	40.6	36.1
SD	14.2	10.3	23.0	2.7	2.2	2.5	10.0	9.4	8.3	20.5	11.8	7.7

correlations relevant to Hypothesis I (PM is a measure of general intellectual functioning) are all high enough to support the hypothesis: PM and WAIS .85; PM and AGCT .70; WAIS and AGCT .84. Inspection of Table 1 also indicates that the intercorrelations do not support Hypothesis II (PM is more significantly related to Cohen's Factor B, Perceptual Organization, than to Cohen's Factors A and C, Verbal Comprehension and Memory or Attention): Raven and Factor A Verbal Comprehension .79; Raven and Factor B, Perceptual Organization .77; Raven and Factor C, Memory or Attention .63. The intercorrelation between age and PM, — .12, in Table 1, offers no clear support for Hypothesis III (Performance on PM will be inversely related to age).

Factor analysis with a sample as small as 39 is perhaps unwarranted. But a centroid analysis and several varimax analyses essentially support the proposition that the part scores of all these tests, and therefore the total scores, even allowing for overlap, are interrelated by a strong factor of general intelligence. PM, of the three measures used, appears most consistently involved by that g factor. A convincing factor analytic demonstration of this conclusion will have to await the accumulation of additional cases.

E. CONCLUSIONS

1. The high intercorrelations between total scores on the PM, the WAIS, and the AGCT, support the contention that each of these tests is a measure of general intellectual functioning, general mental ability, or general intelligence.

2. The relationship between PM and Cohen's Factor B, Perceptual Organization, is not significantly greater than that between PM and the other Cohen factors.

3. This study, not unexpectedly in view of the limited age range and size of the sample, provides little evidence for a negative relationship between scores on PM and age.

REFERENCES

1. BURKE, H. R. Raven's Progressive Matrices: A review and critical evaluation. *J. Genet. Psychol.*, 1958, **93**, 199-228.

2. COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. Consult. Psychol.*, 1957, **21**, 283-290.

3. ————. A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *J. Consult. Psychol.*, 1957, **21**, 451-457.

4. HALL, J. C. Correlation of a modified form of Raven's Progressive Matrices (1938) with the Wechsler Adult Intelligence Scale. *J. Consult. Psychol.*, 1957, **21**, 23-26.

5. RAVEN, J. C. Guide to Using Progressive Matrices (1938). London: H. K. Lewis, 1956.
6. ———. Army General Classification Test, Technical Report. Chicago: Science Research Associates, 1960.
7. WECHSLER, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corp., 1955.

VA Hospital
East Orange, New Jersey 07019

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 211-219.

THE EFFECT OF VERBAL SUGGESTION ON THE REVERSAL RATE OF THE AMES TRAPEZOID ILLUSION* [1]

Department of Psychology, Fordham University

MARGARET KATHLEEN CAPPONE

A. INTRODUCTION

The basic characteristic of the Ames Trapezoid Illusion is the perception of a rotating trapezoidal figure as oscillating back and forth. Since the first description of the phenomenon in 1951 (2), experimental research into the perception of the illusion has centered upon objective factors; wherein, such parameters as viewing distance, rotation speed, figural shape, background intensity, and others have been studied (1, 2, 17, 18, 27, 28, 29, 30, 36, 43, 44, 45). It is apparent from these investigations that the physical properties of the stimulus and other objective factors do act to increase or decrease the perception of the illusion.

Such wide ranges of individual differences have been reported with the Ames phenomenon that intersubject variability has been almost as great as the variability due to altering objective conditions for a given subject. That this differential response on the part of the subject may be due to the participation of external social factors such as suggestion has not been systematically investigated in studies of the Ames Illusion.

Research dealing with suggestion is vast, and many factors have been found to be operative in increasing or decreasing the effectiveness of this variable. Generally, suggestibility is negatively correlated with individual's level of development increasing from birth to age seven or nine after which time it decreases (13, 23, 26, 31, 37). Conflicting evidence exists regarding the relationship between intelligence and suggestibility; some investigators reporting little or no relation (4, 24, 40, 41, 42), while others report a positive cor-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 16, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was submitted in partial fulfillment of the requirements for the degree of Master of Arts in Psychology at Fordham University, Graduate School of Arts and Sciences, Department of Psychology, Bronx, New York. The author wishes to express her sincere gratitude to Joseph G. Keegan, S.J., and Richard T. Zegers, S.J., for their careful guidance and continued inspiration as mentors of this thesis. The development and successful completion of this study was a result of their invaluable help.

relation (16, 19). On the whole, females are slightly more suggestible than males in both children and adults, normals and neurotics (7, 11, 14, 38, 39). However, a few studies report little or no difference between males and females in terms of suggestibility (10, 22); or, that males are more suggestible than females up to age seven after which time females are more suggestible (13). Experimental results pertaining to the problem of whether neurotics are more suggestible than normals have been inconsistent. Several investigators report that neurotic patients are more suggestible (3, 11, 12, 38), while others have reported no differences in suggestibility between neurotics and normals (5, 9, 25).

The structure of the stimulus situation has also been found to be operative in increasing or decreasing the effectiveness of suggestion. In terms of an ambiguous stimulus situation, the general finding has been that suggestion is effective since it establishes a frame of reference for the subject (33, 34, 35). As pointed out by Schonbar (32), high structure interferes with and sometimes precludes the influence of suggestion. The indications from Sherif's (33) work with the autokinetic phenomenon are that the presence of suggestion in an ambiguous situation furnishes a degree of external structure to the stimulus by providing a frame of reference apart from the individual's. Other studies have also reported the modification of perceptual illusions through suggestion (6, 8, 15, 20, 32). Apparently, the effectiveness of suggestion in an unstructured stituation does not lay in changing a previous attitude of the subject; rather does it "redefine" an ambiguous situation (21) and provides a degree of external structure to the situation by contributing a frame of reference to the subject (33).

The primary purpose of the present study was to investigate the effects of individual verbal suggestion upon the reversal rate of the Ames Trapezoid Illusion; and, secondly, to observe the effects of individual verbal suggestion upon the variability of the Ames Illusion. While it is granted that the illusion is an ambiguous situation and as such is limited in providing the subject with objective frames of reference, it would appear that the introduction of suggestion would redefine the ambiguity of the stimulus and provide a degree of external structure by contributing a frame of reference to the subject, other than his own, upon which to make his judgments. Specifically, it was hypothesized that the suggestion, administered individually to each subject, that he should be seeing a greater number of reversals would produce a significant increase in the reversal rate of the Ames Illusion for subjects naive to the experimental situation.

B. APPARATUS AND PROCEDURE

The purpose of this investigation was to determine the quantitative effects of suggestion on the Ames Trapezoid Illusion. The apparatus consisted of one windowed trapezoid, 13.5 inches in length with long and short edges of 13 and 9 inches respectively, mounted on a vertical shaft and driven clockwise by a constant speed motor at a rate of 21 revolutions per minute. The plywood trapezoid figure was coated with luminous paint and was charged before each experimental session by rotating it for 10 minutes under the direct illumination of a 100-watt bulb. This permitted the figure to be viewed in a completely darkened room free from any extraneous background cues.

Five males and five females 14 to 19 years of age served as subjects and all were initially naive with regard to the Ames Illusion. All subjects had normal vision as determined by the Keystone Visual Survey Tests. The visual acuity, depth perception, and lateral and vertical phoria for both near and far usable vision was assessed for each subject. Individuals were accepted as subjects whose performance met the minimum criteria demanded by the test; i.e., received a rating of "accepted" (completely satisfactory performance), or "acceptable" (allowing for slight deficiency). Since there are two possible reversals for every actual rotation of the trapezoid, the maximum number of reversals which could be seen in a three-minute viewing period at 21 rpm was 126. A one-hour screening session which simulated actual experimental conditions (with the absence of suggestion) served to isolate as subjects those individuals who perceived 60 per cent: i.e., 76 reversals, or less, of the maximum number of reversals for three minutes of viewing. These "low-perceivers" were chosen as subjects, since they would not already have perceived the maximum prior to suggestion, but could progress towards the maximum after suggestion.

The procedure was arranged so that each subject was tested separately for 11 consecutive sessions which were spaced at 24-hour intervals and at the same time each day. An individual session consisted of a dark-adaptation period of 10 minutes and 10 test-trials, each of which was three minutes in length with a one-minute rest period between each trial. The subject was to record each perceived reversal on an automatic handcounter. The total length of each session was approximately one hour. The first session was the screening session and also served as the preliminary practice session for each of the 10 subjects finally selected. During this session subjects were given a short instruction on the apparatus used, a description of the reversal phenomenon, and instruction into the nature of their task. Within the experiment proper, the first five sessions constituted the nonsuggestion period and the last five sessions the suggestion period. On the day each was to start his sixth experimental session, the suggestion was given to him individually that he should be perceiving more reversals for his age than he was presently seeing. To reinforce this suggestion the subject's attention was directed to a fake graph contained within a dummy research report which clearly showed an increase in perception as function of age.

C. Results and Discussion

The number of reversals yielded by the 10 three-minute testing periods per day were averaged, and mean and *SD* values obtained for each of the five nonsuggestion and five suggestion sessions for each subject. These data are presented in Table 1. It was hypothesized that suggestion would increase the perception of the illusion. It can be seen from Table 1 that nine of the 10 subjects exhibited the expected rise in mean number of perceived reversals following the administration of suggestion. Subject 10 exhibited a general decrease in his mean score after suggestion was introduced.

In order to study the over-all effects of suggestion upon perception of reversals, the analysis of variance technique was applied to the data. It should be pointed out that the analyses of variance and Tukey tests for the location of the significant differences were computed only on the data of Subjects 1 to 8. The data of two subjects (*S*s 9 and 10) were not included because both subjects were tested together, while all others were tested individually.[2]

The first of these analyses employed subjects and nonsuggestion and suggestion conditions as main sources of variation. The *F* ratio for the main effect of nonsuggestion and suggestion was found to be significant at the .01 level. From this it may be concluded that verbal suggestion designed to increase perception of the illusion effected a positive influence and significantly raised the mean number of reversals perceived during the suggestion period. These findings are in agreement with Berrick (6), Cantril (8), Kelly (20), Schonbar (32), and Sherif (33), all of whom found that perceptual illusions could be modified through the use of suggestion.

Two other analyses of variance were computed in order to ascertain the presence of a general practice effect. The first of these analyses was computed with the mean reversal rates of the five nonsuggestion sessions and the subjects serving as main effects. The resulting *F* ratio did not attain sig-

---

[2] Subjects 9 and 10 were tested together in order to complete all the sessions; however, since this may have influenced their data it was not included in the statistical analyses.

TABLE 1

THE MEAN REVERSAL RATES AND STANDARD DEVIATIONS PER 3 MINUTES FOR 10 SUBJECTS FOR THE NONSUGGESTION AND THE SUGGESTION SESSIONS

Nonsuggestion sessions

	1		*2*		*3*		*4*		*5*	
Subjects	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	81.2	9.64	87.9	10.86	83.6	7.14	75.9	11.14	81.3	11.53
2	51.7	7.07	52.8	13.53	59.6	7.87	66.6	9.59	60.5	11.14
3	39.1	6.33	45.5	8.87	39.2	12.51	32.0	18.41	45.9	7.81
4	33.2	10.39	42.2	4.44	41.7	4.78	46.1	6.12	49.0	8.69
5	34.2	23.71	31.1	16.41	53.0	23.52	32.4	18.03	41.4	24.11
6	8.7	7.41	3.6	3.63	1.4	4.27	5.3	3.53	4.6	7.68
7	34.6	8.22	42.2	7.67	42.0	10.79	51.5	7.08	48.1	3.19
8	59.2	17.72	79.2	7.55	61.3	23.26	42.2	14.25	44.7	12.69
9	1.8	7.42	1.5	8.85	0.0	0.00	0.0	0.00	0.0	0.00
10	23.6	18.71	28.9	15.64	24.8	7.89	33.1	19.85	20.2	11.87

Suggestion sessions

	6		*7*		*8*		*9*		*10*	
1	122.4	1.00	120.9	4.24	123.9	2.24	118.1	7.35	120.0	2.64
2	98.3	5.29	97.8	7.35	99.4	6.16	111.8	6.00	99.7	5.75
3	100.6	27.15	118.4	9.21	112.5	5.97	119.1	3.96	121.5	3.75
4	100.3	10.75	111.9	4.97	116.0	4.62	118.0	4.37	115.9	2.96
5	85.8	10.54	92.9	19.18	105.4	5.18	105.2	10.68	104.8	9.42
6	14.3	10.95	15.8	5.52	15.4	2.63	14.8	2.31	15.3	2.45
7	113.2	15.49	116.7	4.42	118.6	4.70	113.9	5.31	120.1	3.16
8	61.2	12.00	96.4	14.04	96.6	11.40	90.0	13.19	122.5	3.10
9	45.5	11.96	32.5	9.89	42.7	17.88	41.8	14.07	42.9	16.25
10	12.2	8.43	22.4	3.23	18.4	7.71	35.1	7.68	9.0	7.21

nificance and indicated that there was no increase in perceived reversals among nonsuggestion sessions. That is, that continued exposure to the stimulus situation did not, in and of itself, result in an increase in the perception of the illusion. This finding is in agreement with Murray (29) who found learning to be completed during the first hour of viewing. The second analysis of variance was performed upon the suggestion session means, with subjects and sessions as main effects. The sessions constituting the suggestion period were found to be significantly different from each other at the .05 level. The Tukey procedure was applied to this data and it was found that significantly more reversals were perceived by the subjects as a group during Sessions 8, 9, and 10 than there were in Session 6 of the suggestion series. The implications here are that the effects of an initial suggestion given once to each subject (at the start of his sixth session) continued to increase throughout succeeding sessions without the presence of reinforcement from the experimenter.

In order to ascertain the overall effect of suggestion upon the variability of the subjects in the perception of the illusion, the analysis of variance technique was applied to the standard deviation values of each subject, taken over all sessions of nonsuggestion and suggestion. The main sources of variation in the first analysis were subjects and nonsuggestion and suggestion. It was found that there was a significant difference between the nonsuggestion and suggestion sessions in terms of their standard deviations. On this basis it was concluded that the verbal suggestion utilized in the present study acted to decrease significantly the variability of the subjects. This decrease in variability may be explained on the basis that verbal suggestion provided the subject with a frame of reference in an objectively ambiguous situation.

Two separate analyses of variance were computed in order to investigate whether continued observation had any effect upon variability. The nonsuggestion and suggestion *SD*'s did not yield significant *F* ratios and, therefore, no practice effect was found to occur for either of these periods.

Generally, the results of this research indicate that verbal suggestion designed to increase the perception of the Ames Illusion resulted in significantly increasing the mean reversal rate and decreasing the variability of the subjects. In terms of the *SD* values for both the nonsuggestion and suggestion sessions, no practice effect was found. Neither was there a practice effect evident in the $\overline{M}$ reversal rate values for the nonsuggestion period. (It seems reasonable to infer that had suggestion *not* been administered to the subjects in Sessions 6 to 10 they would have continued perceiving the illusion at the same reversal rate as in Sessions 1 to 5. The basis for this inference lies in the fact that there was no significant difference between any of the nonsuggestion sessions in the $\overline{M}$ number of illusions perceived by the eight subjects.) However, within the suggestion trials there was a significant increase in mean values, with the last suggestion session being significantly larger than the first suggestion session; this was interpreted as indicating the continual influence of an initial suggestion. In conclusion, therefore, the results substantiate the view posited by Sherif (33) that suggestion apparently operates as a modifying power in an ambiguous stimulus situation by creating an objective, external frame of reference on which the subject can base his judgments. In view of the results of this experiment, it appears warranted to conclude that, within the limits imposed by the research design, verbal suggestion administered individually will cause an alteration in the reversal rate of the Ames Illusion.

D. Summary and Conclusion

In order to determine the influence of verbal suggestion in effecting a significant increase in the perception of the Ames Trapezoid Illusion, the present investigation was undertaken. Five males and five females served as subjects, each being tested separately for 10 consecutive experimental sessions; the first five sessions constituted the nonsuggestion period, and the last five sessions the suggestion period.

The results discussed in the report may be summarized as follows:

1. Individual verbal suggestion significantly increased the perception of the Ames Illusion at the .01 level for the eight subjects whose data was treated statistically.
2. For this group of subjects there was no practice effect evident.
3. There appeared to be a tendency for the effectiveness of suggestion to increase over the five suggestion sessions.
4. There was a significant decrease in variability following the introduction of suggestion.

References

1. Allport, G. W., & Pettigrew, T. F. Cultural influence on the perception of movement: The trapezoidal illusion among the Zulus. *J. Abn. & Soc. Psychol.*, 1957, **55**, 104-113.
2. Ames, A. Visual perception and the rotating trapezoidal window. *Psychol. Monog.*, 1951, **65**, No. 7 (Whole No. 324).
3. Arcieri, L. Differences in the degree of suggestibility between schizophrenics, psychoneurotics, and normal subjects in the reproduction of forms. *Psychiat. Quart.*, 1949, **23**, 41-58.
4. Aveling, F., & Hargreaves, H. L. Suggestibility with and without prestige in children. *Brit. J. Psychol.*, 1921, **12**, 53-65.
5. Bartlett, M. R. Suggestibility in psychopathic individuals: A study with psychoneurotic and dementia praecox subjects. *J. Gen. Psychol.*, 1936, **14**, 241-247.

6. Berrick, M. An investigation of the influence of examiner suggestion on the perception of ambiguous stimuli. *Diss. Abst.*, 1955, **15**, 2566.
7. Brown, W. Individual and sex differences in suggestibility. *Univ. Calif. Publ. Psychol.*, 1916, **2**, 291-430.
8. Cantril, H. The prediction of social events. *J. Abn. & Soc. Psychol.*, 1938, **33**, 364-389.
9. Davis, L. W., & Husband, R. W. A study of hypnotic susceptibility in relation to personality traits. *J. Abn. & Soc. Psychol.*, 1931, **51**, 634-638.
10. Edwards, A. S. An experimental study of sensory suggestion. *Amer. J. Psychol.*, 1915, **26**, 99-124.
11. Eysenck, H. J. States of high suggestibility and the neuroses. *Amer. J. Psychol.*, 1944, **57**, 406-411.
12. Furneaux, W. D. The prediction of susceptibility to hypnosis. *J. Personal.*, 1946, **14**, 281-294.
13. Gilbert, J. A. Researches on the mental and physical development of school children. *Stud. Yale Psychol. Lab.*, 1894, **2**, 40-100.
14. Heron, W. T. The group demonstration of illusory warmth as illustrative of the phenomenon of suggestion. *J. Abn. & Soc. Psychol.*, 1927, **22**, 341-344.
15. Hopkin, V. D. Some effects of influence on the perception of an illusion in small groups. *Percept. & Motor Skills*, 1957, **7**, 212.
16. Hull, C. L. Hypnosis and Suggestibility. New York: Appleton-Century, 1933.
17. Ittelson, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285-294.
18. Ittelson, W. H., & Kilpatrick, F. P. Experiments in perception. *Sci. Amer.*, 1951, **185**, 50-55.
19. Jones, G. M. Experiments upon the reproduction of distances as influenced by suggestions of ability and inability. *Psychol. Rev.*, 1910, **17**, 269-278.
20. Kelly, E. L. The effect of previous experience and suggestion on the perception of apparent movement. *Psychol. Bull.*, 1935, **32**, 569-570.
21. Lewis, H. B. Studies in the principles of judgments and attitudes: IV. The operation of "prestige suggestion." *J. Soc. Psychol.*, 1941, **14**, 229-256.
22. Lindberg, B. J. Suggestibility in different personality types. *Amer. J. Psychol.*, 1940, **53**, 99-108.
23. Lodge, J. H. The illusion of warmth test for suggestibility. *Forum Educ.*, 1926, **40**, 180-186.
24. McGeoch, J. A. The relationship between suggestibility and intelligence in delinquents. *Psychol. Clin.*, 1925, **16**, 133-134.
25. Messer, A. L., Hinckley, E. D., & Mosier, C. I. Suggestibility and neurotic symptoms in normal subjects. *J. Gen. Psychol.*, 1938, **19**, 391-399.
26. Messerschmidt, R. The suggestibility of boys and girls between the ages of six and sixteen years. *J. Genet. Psychol.*, 1933, **43**, 422-437.
27. Mulholland, T. B. Motion perceived while viewing rotating stimulus objects. *Amer. J. Psychol.*, 1956, **69**, 96-99.
28. ———. The "swinging disk" illusion. *Amer. J. Psychol.*, 1958, **71**, 375-382.
29. Murray, P. A. A quantitative study of the Ames Oscillatory Illusion. Unpublished Doctoral dissertation, Fordham University, New York, 1961.
30. Pastore, N. Some remarks on the Ames oscillatory effect. *Psychol. Rev.*, 1952, **59**, 319-323.
31. Reymert, M. L., & Kohn, H. A. An objective investigation of suggestibility. *Psychol. Bull.*, 1938, **35**, 683-684.
32. Schonbar, R. A. The interaction of observer-pairs in judging visual extent and movement. *Arch. Psychol.*, 1945, No. 299.

33. Sherif, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.

34. ———. The Psychology of Social Norms. New York: Harper, 1936.

35. Sherif, M., & Sherif, C. W. An Outline of Social Psychology. New York: Harper, 1956.

36. Slack, C. W. Critique on the interpretation of cultural differences in the perception of motion in Ames' trapezoidal window. *Amer. J. Psychol.*, 1959, **72**, 127-131.

37. Small, M. H. Suggestibility in school children. *Ped. Sem.*, 1896, **13**, 176-220.

38. Stukat, K. G. Suggestibility: A Factorial and Experimental Analysis. Stockholm: Almquist & Wiksell, 1958.

39. Tuddenham, R. D. The influence of a distorted group norm upon individual judgment. *J. of Psychol.*, 1958, **46**, 227-241.

40. Wegrocki, H. J. The effect of prestige suggestibility on emotional attitudes. *J. Soc. Psychol.*, 1934, **5**, 384-394.

41. White, R. S. The influence of suggestibility on responses in an ink blot test. *Child Devel.*, 1931, **2**, 76-79.

42. Young, P. C. Intelligence and suggestibility in whites and negroes. *J. Comp. Psychol.*, 1929, **9**, 339-359.

43. Zegers, R. T. The reversal illusion of the Ames trapezoid. *Trans. N. Y. Acad. Sci.*, 1964, **26**, 377-400.

44. Zegers, R. T., & Murray, P. A. Perception of distortion: An experimental approach to illusion. Technical Report Navtradevcen, U.S. Naval Training Device Center, New York, 1962, 377-400.

45. Zuckerman, B. C. The phenomenal movement of the rotating trapezoidal window; an experimental investigation. Unpublished Master's dissertation, New School for Social Research, New York City, 1953.

665 Thwaites Place (6A)
Bronx 67, New York

Published as a separate and in *The Journal of Psychology*, 1966, 62, 221-228.

THE RELATION OF VERBAL INTELLIGENCE AND COLOR-FORM DISCRIMINATIVE ABILITY TO CHILDREN'S COLOR-MATCHING AND FORM-MATCHING BEHAVIOR*

State University of New York at Buffalo, and Washington University, St. Louis

NORMAN L. CORAH, SALLY ANN JONES, AND BARBARA B. MILLER[1]

A. PROBLEM

Studies conducted in the last several decades have demonstrated that, when asked to match geometric shapes that are comparable in color or form, children between the ages of three and six years use color more often than do older children. The present report is part of a program of research designed to investigate the variables that influence color matching in children's color-form perception (3, 4, 5).

Several investigators have suggested that form matching is related to higher intelligence. As early as 1913, Katz (12) attributed superior intellectual development to the few form-dominant subjects among the children in his study. Tobie (17) also expressed the view that form matching increases with the development of intelligence. Engel (7) divided his school children into groups of good, moderate, and poor "intelligence" and found the highest percentage of form responses in the first group. The highest percentage of color responses was obtained from the "poor" group. Lindberg (13) found that giving initial color responses on his Ring abstraction test was associated with poor school achievement in children. However, he found no relationship between color-form sorting and school achievement. Both Brian and Goodenough (1) and Colby and Robertson (2) suggested that form matching was associated with high mental-test scores, but neither study provided any data in support of this contention.

Honkavaara (11) has presented data that suggest that *color* matching is

---

* Received in the Editorial Office, Provincetown, Massachusetts, on February 18, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Parts of this report were adapted from the Senior Honor's Thesis of the second author and the Master's Thesis of the third author. The writers are indebted to the administrators and staffs of the following St. Louis area schools for providing subjects: Bel-Nor and Bel-Ridge Normandy Public Schools, Community School, Ladue Chapel Nursery School, and Washington University Nursery School. The data analysis was facilitated by support of the Washington University Computer facilities through National Science Foundation Grant G-22296.

related to high intelligence. However, her statistics were only descriptive and the intellectual groupings of her subjects were inferred from school background. While there has been much discussion of the relationship between color-form perception and intelligence, there is a dearth of data in terms of measured intelligence. The present report attempts to explore the hypothesized relationship between form matching and intelligence.

This research also explores the relationship between color-form matching and color-form discriminative ability of children. Gaines (9, 10) has presented data that indicate that a stylistic component is involved in color-form perception. She found that color responders discriminated color better than form, while form responders discriminated form better than color.

Our previous work (3, 4) has proceeded from the assumption that the attentional process involved in color matching in young children is analogous to Piaget's (14) process of centration: i.e., the child's attention is centered upon the dominant color characteristic of the stimulus configuration. As the child's perception becomes decentered, he attends more readily to form. Since decentered perception is assumed to be a more highly developed genetic process than centered perception, it would be expected that children who typically decenter their perception will demonstrate greater discriminative ability in all areas of perception than children who typically center their perception. Consequently, it would be expected that form responders will discriminate both color and form more accurately than will color responders. It should also be noted that this approach leads to a prediction of no relationship between intelligence and color-form matching because Piaget does not view centration effects as being related to intelligence (8).

B. Study I

1. *Method*

a. Subjects. The subjects (*S*s) came from three nursery schools in the St. Louis area. There were 48 children, 24 boys and 24 girls, ranging in age from 45 to 65 months with a mean age of 56.94 months.

b. Materials. The color-form test consisted of 18 stimulus plates on which a color match or a form match could be made. Each standard figure was located at the top of an 11 × 8-1/2 inch sheet of white paper. Two comparison figures were placed, one at the right and one at the left, below the standard. One of the comparison figures matched the standard in color but not in form, while the other matched it in form but not in color.

Three basic standard figures and three colors—red, blue, and green—were used. The figures were asymmetric forms with five to seven sides. The comparison figure of the same color as the standard retained the same general configuration but was readily discriminable from the standard in shape. An example of these figures has been given elsewhere (4). The three basic figure plates were presented six times in the different colors. The score used was the total number of color responses out of a possible 18. The reliability of this color-form test is high. Test-retest data, obtained over an interval of three to four weeks, from 40 nursery school children and 40 second graders yielded correlations of .91 and .93 respectively.

The Peabody Picture Vocabulary Test, Form A (6) was used as a measure of intelligence. The color-discrimination test consisted of 20 3 × 5 inch cards with four Munsell color chips pasted on each. One of the four color chips on each card differed from the other three in degree of saturation. These differences in degree of saturation became progressively more difficult to discriminate as the test proceeded. The highest possible score was 20 correct choices. Three additional simple practice cards were included but not scored.

The form-discrimination test was similar to the color-discrimination test in format. Four figures were drawn on a 5 × 8 inch card. One varied from the other three in terms of internal angle size. Four basic figure designs were used. Each was presented six times. Differences between the one "different" and three identical figures on each card were progressively more difficult to discriminate. The highest possible score was 24 correct choices. In addition, three simple pretrial practice cards were used.

c. Procedure. The color-form test was administered to 133 preschoolers. As each plate was presented, the examiner said: "Which one of these looks most like this one?" As she said "this one," she pointed to the standard at the top of the page. Children who clearly did not understand the task or who showed side preferences, defined as six or more consecutive responses to the same side, were excluded. As a result, 21 *S*s were excluded by these criteria, leaving a total of 112.

Form responders and color responders were designated as those *S*s in the lower third and upper third respectively of the score distribution. That is, form responders were designated as those with scores of zero to 5 and color responders as those with scores of 13 to 18. This procedure yielded 24 *S*s, 12 boys and 12 girls, in each of the groups. Statistical analysis revealed no differences in age between the groups.

The intelligence test was administered in a second session and the two discrimination tests in a third session with each child. In presenting each item of the two discrimination tests, the examiner asked: "Which one of these is *not* like the others?"

2. *Results*

The score distributions were either markedly skewed or bimodal; consequently, nonparametric statistics were used for analysis. The results are presented in Table 1.

TABLE 1
INTELLIGENCE, COLOR, AND FORM DISCRIMINATION SCORES FOR THE TWO GROUPS

	Form responders $N = 24$		Color responders $N = 24$		
Measure	Range	Median	Range	Median	U
Intelligence	92-127	113.0	53-129	112.5	280.5
Color discrimination	0-17	14.5	6-16	13.0	233.5
Form discrimination	1-17	7.5	0-12	7.5	231.0

The Mann-Whitney *U* test (16) was used to assess differences between the groups. None of the differences was significant. Sex differences were also assessed and found to be not significant.

C. STUDY II

1. *Rationale and Method*

Since some of our previous work has suggested that preschool *S*s give less reliable responses to the measures employed in these studies, the study was replicated with an older group of children. In addition, the color-form test was revised to provide greater latitude of response.

The color-form test included simple geometric forms—square, circle, triangle—as well as some of the asymmetric figures used in the first study. An additional color, yellow, was added to the series. This revised test consisted of 20 plates. The other tests were the same as in the previous study.

The color-form test was administered to 227 kindergarten children from two public schools. Forty *S*s were excluded for giving side preferences. A color responder was designated as a child who gave 16 or more color responses, while a form responder was designated as a child who gave 16 or more form responses. This procedure yielded many more form responders than color responders. The form responders were randomly sampled with the restriction that their age range approximate that of the color responders.

As a consequence, two groups of 24 *S*s, 12 boys and 12 girls each, were obtained. The 48 *S*s ranged in age from 62 to 81 months with a mean age of 67 months. The mean age of each group was the same as that for all *S*s.

2. *Results*

Two of the score distributions again departed from normality. Consequently, the statistical analyses used in the first study were again applied. The results are given in Table 2.

TABLE 2
INTELLIGENCE, COLOR, AND FORM DISCRIMINATION SCORES FOR THE TWO KINDERGARTEN GROUPS

	Form responders		Color responders		
	$N = 24$		$N = 24$		
Measure	Range	Median	Range	Median	U
Intelligence	74-145	114.5	78-132	109.0	200.5*
Color discrimination	8-18	14.0	8-15	13.0	178.5**
Form discrimination	0-20	11.5	0-22	11.0	249.0

* $p = .07$.
** $p < .05$.

There was a tendency for the form responders to have higher *IQ*s than the color responders, although the difference does not reach the .05 level of significance. The form responders were able to discriminate color significantly better than the color responders. The difference in form discrimination was not significant. No significant sex differences were found.

D. Study III

Another study conducted in our laboratories adds some additional data on the relationship between color-form matching and intelligence. One of the groups of *S*s in this study was composed of 40 third-grade children, 20 boys and 20 girls, who ranged in age from 104 to 117 months with a mean age of 108 months.

The *S*s were given the 18-plate color-form test used in the first study.[2] All of the *S*s had been given the Stanford-Binet Intelligence Test (Form L-M) at school. A Spearman rank-difference correlation was calculated between the number of color responses and *IQ* for the 40 *S*s. A rho of $-.304$ ($p = .07$) was obtained. There is a marked tendency for form matching to be associated with higher intelligence, although the relationship again falls short of significance at the .05 level.

E. Discussion

In two of the three age groups studied, a relationship between form matching and higher intelligence was suggested. However, it would seem that, in the present context, the question of practical significance would have to be raised.

---

[2] The authors are indebted to Eva Gospodinoff who collected the data and made the results available.

Many of the early writers attributed form matching, especially in younger children, to the assumed higher intelligence of such children. When the present data are viewed in terms of the extreme color-form response groups selected for study, it must be concluded that intelligence plays no marked role in color-form matching behavior.

It would appear that the same conclusion must be made with regard to the hypotheses concerning color and form discrimination. However, these data raise some interesting questions. In studies I and II, the form-discrimination scores were in the predicted direction but not significantly different. The color-discrimination scores were also in the predicted direction and the difference was significant in the second study. These results are directly opposite to those predicted and obtained by Gaines (9, 10) in her color-form work with older children.

In previous work (3, 4), it has been suggested that a dual basis exists for color-matching behavior. A color match may result from an attentional centration effect or from a learned preference for color. It has been suggested that the color matching of younger children is primarily of the first type, while that of older children is, for the most part, preferential. The lack of any difference in intelligence in the preschool groups of Study I would tend to support this idea since centration effects and intelligence should be uncorrelated (8, 15). However, it is possible that color matching as a preference response may be associated with lower intelligence. In that case, the conclusions made above would have to be tempered somewhat.

This series of studies (3, 4, 5) has tended to use somewhat complex stimuli which might emphasize the centration effect in younger children. Gaines (9, 10) and others have used less complex, and possibly more meaningful, stimuli with which color matching might be almost entirely a preference response. These differences in stimuli might explain the different results which have been obtained with children of the same age. Further research is needed to clarify these issues.

F. SUMMARY

The research in this report was concerned with the relationship between three variables—intelligence, color discrimination, and form discrimination—and color-form matching behavior in children. While early writers predicted an inverse relationship between color matching and intelligence, the present approach led to a prediction of no relationship. It was further predicted that form responders would discriminate both color and form more accurately than color responders.

A group of 24 preschool color responders and a group of 24 preschool form responders were selected with a color-form matching test. Measures of intelligence, color-discrimination, and form-discriminative ability were obtained. No significant differences were found between the two groups on the dependent variables. The study was replicated with two groups of 24 color responders and form responders of kindergarten age. Form responders had higher *IQ*s and form-discrimination scores than did color responders, but the differences fell short of significance. Form responders discriminated color with significantly greater accuracy than color responders. Intelligence and color-form response were also assessed in a group of 40 third graders. Again, *IQ* was related to form matching, but the relationship fell short of statistical significance.

A dual basis for color matching was postulated. Color response can be considered as a centration effect or as a learned preference. The differences in the obtained results, as well as those in other studies, were discussed in relation to these two processes.

REFERENCES

1. BRIAN, C. R., & GOODENOUGH, F. L. The relative potency of color and form perception at different ages. *J. Exper. Psychol.,* 1929, **12**, 197-213.
2. COLBY, M. C., & ROBERTSON, J. B. Genetic studies in abstraction. *J. Comp. Psychol.,* 1942, **33**, 385-401.
3. CORAH, N. L. Color and form in children's perceptual behavior. *Percept. & Motor Skills,* 1964, **18**, 313-316.
4. ———. The effect of some stimulus characteristics on color-form perception in preschool children. Paper read at the Biennial Meeting of the Society for Research in Child Development in Minneapolis, Minnesota, March 25, 1965.
5. ———. The effect of instruction and performance set on color-form perception in young children. *J. Genet. Psychol.,* 1966, **108**, in press.
6. DUNN, L. M. Peabody Picture Vocabulary Test. Philadelphia: American Guidance Service, 1959.
7. ENGEL, P. Uber die teilinhaltliche Beachtung von Farbe und Form. Untersuchung an 800 Schulkindern. *Z. F. Padag. Psychol.,* 1935, **36**, 202-214, 241-251.
8. FLAVELL, J. The Developmental Psychology of Jean Piaget. Princeton: Van Nostrand, 1963.
9. GAINES, R. Color-form preferences and color-form discriminative ability of deaf and hearing children. *Percept. & Motor Skills,* 1964, **18**, 70.
10. ———. Color-form preference and color-form discriminative ability of deaf adolescents. Paper read at Midwestern Psychological Association Meeting in St. Louis, May 1, 1964.
11. HONKAVAARA, S. A critical reevaluation of the color and form reaction, and disproving of the hypotheses connected with it. *J. of Psychol.,* 1958, **45**, 25-36.
12. KATZ, D. Studien zur Kinderpsychologie. *Wiss. Beitr. Padag. u. Psychol.,* 1913, **4**.

13. LINDBERG, B. J. Experimental studies of colour and noncolour attitude in school children and adults. *Acta Psychiat. & Neurol.*, 1938, Suppl. 16.

14. PIAGET, J. The Psychology of Intelligence. London: Routledge & Kegan Paul, 1950.

15. POLLACK, R. H. Simultaneous and successive presentation of elements of the Mueller-Lyer figure and chronological age. *Percept. & Motor Skills*, 1964, **19**, 303-310.

16. SIEGEL, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

17. TOBIE, H. Die Entwicklung der teilinhaltlichen Beachtung von Farbe und Form im vorschulpflictigen Alter. *Z. f. Angew. Psychol.*, 1926, **38**.

State University of New York at Buffalo
The University Health Center
Capen Hall
Buffalo, New York 14214

Published as a separate and in *The Journal of Psychology,* 1966, **62**, 229-240.

PERSONALITY CHANGES IN PSYCHIATRIC RESIDENTS DURING TRAINING*

Veterans Administration Hospital, Seattle; University of Washington; and Western Research Support Center, Veterans Administration Hospital, Sepulveda, California

WILLIAM S. KOGAN,[1] ERLING E. BOE,[2] EDWARD F. GOCKA,
AND MERLIN H. JOHNSON[1]

A. INTRODUCTION

A problem of considerable significance in the study of personality is the assessment of changes that apparently occur as a function of major life events (e.g., beginning school, getting married, undertaking psychotherapy, etc.). The professional training of psychiatrists constitutes one specific instance of such events whose influence has been observed and reported in qualitative terms (7). In the following summary statement, Holt describes the personality changes in a group of psychiatric residents that became apparent to him during the course of their training:

> I have tried to show how psychiatric training itself contributes to maturity through the deepening of interests, the development of self-insight, the development of sound relationships with people, and the stabilizing of ego identity. At the same time, the training process induces apparently regressive changes, seen in morale problems and the loss of the initial enthusiastic glow (7, p. 214).

Although qualitative observations such as these are interesting and valuable insofar as they go, a number of important questions about the nature of personality change can not be answered adequately by this method. One problem is the identification of relatively independent dimensions on which change occurs. Many changes might be too subtle for detection by unaided human observation. Furthermore, some dimensions on which changes are detected might be so highly correlated that they could be combined without loss of

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 2, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Also affiliated with the University of Washington School of Medicine.

[2] This research was accomplished while the second author was at the Veterans Administration Hospital, Seattle.

information. A second problem involves precise measurement of the degree to which change occurs. No doubt some changes are greater than others. A third major problem is the determination of rate of change. The traditional strategy of making observations before and after some event is obviously inadequate for plotting change as a function of time. To plot this function, a series of observations systematically spaced over time are required.

One purpose of the present research was to provide a systematic and quantitative analysis of some personality changes that occur in psychiatric residents during the course of their training. It was specifically designed so that significant dimensions of change could be identified, amount of change measured, and rate of change plotted. Because no previous quantitative research on personality changes in psychiatric residents has been reported, the present effort was intended to be exploratory. It was judged appropriate to begin the search for significant dimensions of change within the general area of interpersonal variables. Since psychiatric training primarily involves interpersonal relations among resident, patient, and supervisor, it seems reasonable to expect that at least some of the residents' perceptions with regard to interpersonal behavior would undergo change.

Measurement of interpersonal variables was accomplished by administering Kogan and Fordyce's (9) revision of the Interpersonal Check List (11) on five occasions during training. However, scores on the predetermined eight scales of this instrument were not used because previous research (10) indicated that they tended to be insensitive to change. Instead, the 64 items of the Revised Interpersonal Check List (R-ICL) were treated as a reservoir within which significant dimensions of change would be sought. As with other techniques for the assessment of interpersonal behavior (3), the R-ICL has a circumplex structure around two orthogonal axes of dominance-submissiveness and love-hate. Thus, the items were not completely heterogeneous, but possessed a definite order from the outset.

At a recent conference (6) on problems in measuring change, Horst (8) proposed a factor-analytic solution to the problem of assessing changes inherent in repeated (and multiple) measures over time. Since this method was particularly suited to the analysis of performance on the 64 separate items of the R-ICL, individual item responses were factor analyzed accordingly. This experiment represents an initial attempt to analyze data by Horst's method, and its second major purpose was to apply the procedures resulting from his theoretical work to empirical data in an effort to examine their usefulness in delineating significant dimensions of personality change.

B. Method

1. *Subjects*

The initial pool of *S*s was 19 psychiatric residents with the Department of Psychiatry, University of Washington School of Medicine. These 19 residents began training either in 1959 or 1960. However, because the research entailed repeated psychological testing during the three-year period of the residency, two of the original 19 *S*s were lost—one due to separation from the training program and one because of failure to complete all the tests. Thus, the sample actually used was reduced to 17 *S*s.

All of the remaining 17 *S*s were male and ranged in age from 25 to 50 years with a median of 31 years at the beginning of training. Seven of the residents had previous experience as general practitioners for a range of five to 20 years. The remaining 10 residents began psychiatric training upon completing internship. In general, the applicants admitted to the program were above average academically, exhibited some psychological orientation, and were free of obvious disabling psychopathology.

2. *Test Materials*

All testing was accomplished with the 64 items of the R-ICL (9). The purpose of the revision was to balance each of the eight scales for social desirability (SD) (2). This was done by selecting eight items for each scale with the following characteristics: (*a*) four items had socially desirable scale values and four had socially undesirable scale values, (*b*) the mean SD scale value of the eight-item groups was very near 5.00 (neutral SD) on Edwards' SD rating scale, and (*c*) the standard deviations of the SD scale values of each group of items were nearly equal.

3. *Experimental Design*

A treatments by treatments by subjects design (12, p. 237 *ff*.) was used in which every treatment was administered to all *S*s in the same order. The treatments were five occasions on which the R-ICL was administered and eight instructional sets. Testing was carried out as follows during the course of the three-year residency: zero months, six months, 12 months, 24 months, and 36 months. On every occasion that the R-ICL was administered, each resident responded to all 64 items eight times in terms of the following eight instructional sets (in the order presented): (*a*) standard SD ratings on Edwards' (2) scale, (*b*) Most people are, (*c*) I am, (*d*) The ideal patient is, (*e*) Most patients are, (*f*) As a therapist I am, (*g*) Most therapists are, and (*h*) Ideally, a therapist should be.

4. *Training Program*

The training faculty consisted of a small number of full-time staff (some of whom had psychoanalytic backgrounds) and a larger number of part-time staff. The part-time faculty, who were also in private practice, served as psychotherapy supervisors and as consultants in a more general sense. The program was psychoanalytically oriented with emphasis on the dynamic interrelation of patient, illness, and background. The greatest stress was on the understanding of the patient as a means of helping him deal with his illness and environmental demands.

5. *Procedure*

The senior author usually administered the R-ICL either individually or to small groups of residents. The general outline and objectives of the research project were explained to the residents on the first testing occasion, and instructions specific to their task were presented orally. When unavoidable, they sometimes worked privately at times of their convenience on the later administrations. One instructional set was printed at the top of each answer sheet.

6. *Statistical Analyses*

As noted in the introduction, a decision was made to abandon the original scale structure of the test and to secure independent factors within the reservoir of 64 items. The following steps were followed in statistically analyzing the results: (*a*) total test performance was factor analyzed by item and rotated, (*b*) factor scores were computed, and (*c*) a separate analysis of variance on the factor scores was made for each factor. These three steps will be described more fully below.

The raw data were first arranged in a matrix according to cell M_{14} of Horst's [(8), Formula 39]. This matrix entailed 680 rows (five occasions × eight instructional sets on each occasion × 17 *S*s responding to each set on all occasions) and 64 columns (one for each item). The 64 items were intercorrelated across the 680 rows. The resulting 64 × 64 correlation matrix with unity inserted in the cells of the major diagonal was factor analyzed by the Jacobi method. Then the factors were rotated by the orthogonal varimax method of Kaiser. These and subsequent analyses were performed with an IBM Model 7094 computer.

Factor scores (f) next were obtained through a procedure outlined by Harman [(5), Formula 16.12]. A factor score was thus obtained for every combination of occasion, set, and *S*. The factor score matrix had the same number of rows (680) as the raw data matrix, but fewer columns because items had been combined into independent factors. Hence, the operations were essentially that of data reduction: i.e., individual item responses (true/false scored one or zero) were reduced in number and transformed to factor scores, but none of the two dimensions (occasions, sets) or individual *S*s under which the measures were taken originally were "lost" or changed.

The final step in the statistical analyses was to examine differences in factor scores (for each factor separately) as a function of occasion and instructional set. Analysis of variance was performed by first obtaining a complete factorial breakdown of the data and then recombining for appropriate error terms and nested equivalents in a manner outlined by Gocka (4).

C. RESULTS

The factor analysis yielded 18 orthogonal factors which accounted for 61.1 per cent of the total variance. Seventeen factors were selected which met the arbitrary criterion of roots larger than one. An additional factor was selected because its root was only slightly below one and because such a procedure is consistent with Thurstone's recommendation of rotating one more factor than that given by some criterion stop. Furthermore, the inclusion of factors beyond XVIII substantially altered the loading pattern of Factor XVIII upon rotation. The rotated factor matrix and 64 R-ICL items have been deposited at the American Documentation Institute.[3] Only items with loadings of .40 or greater (except in two instances to improve factor description) and the percentage of total variance accounted for by a factor are presented on page 234.

Rotated factor loadings were then used to compute factor scores. Next, the factor scores were submitted to an analysis of variance for each of the 18 factors separately in order to determine whether the instructional sets produced significant differences (main effect of set), whether scores changed significantly during the five administrations over the three-year residency (main effect of occasion), and whether the occasion and set dimensions interacted. Thus, a total of 54 tests of significance were made for the 18 factors. The *F* ratios obtained from these tests are reported in Table 1. In brief, the main effect of instructional set was significant at the .01 level or beyond for 12 factors, the main effect of occasion was significant at the .01 level or beyond for three factors, but none of the set by occasion interactions reached

---

[3] Order Document No. 8794 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Factor I: 8.61 per cent	
Self-confident	.55
Independent	.53
Respected by others	.43
Often unfriendly	—.45
Irritable	—.64
Shy	—.64
Frequently disappointed	—.65
Resentful	—.66
Lacks self-confidence	—.70
Tender and easily hurt	—.72
Factor II: 3.84 per cent	
Likes everybody	.71
Fond of everyone	.68
Forgives anything	.64
Friendly all the time	.53
Factor III: 3.42 per cent	
Dominating	.67
Acts important	.65
Boastful	.54
Proud and self-satisfied	.49
Expects everyone to admire him	.42
Manages others	.40
Factor IV: 4.55 per cent	
Admires and imitates others	.72
Often helped by others	.66
Eager to get along with others	.55
Very anxious to be approved of	.52
Very respectful of authority	.43
Factor V: 3.80 per cent	
Oversympathetic	.78
Overprotective of others	.70
Too lenient with others	.69
Spoils people with kindness	.53
Factor VI: 2.30 per cent	
Resents being bossed	.66
Factor VII: 2.99 per cent	
Hard to impress	.63
Shrewd and calculating	.61

Factor VIII: 2.15 per cent	
Modest	.79
Appreciative	.38
Factor IX: 5.32 per cent	
Gives freely of self	.67
Good leader	.62
Forceful	.60
Kind and reassuring	.49
Factor X: 2.17 per cent	
Sarcastic	.65
Able to criticize self	.37
Appreciative	.36
Factor XI: 3.06 per cent	
Spineless	.69
Will confide in anyone	.67
Clinging vine	.55
Factor XII: 4.75 per cent	
Can be strict if necessary	.75
Can be frank and honest	.68
Hard-boiled when necessary	.62
Respected by others	.40
Factor XIII: 3.00 per cent	
Often admired	.58
Likes to compete with others	.51
Respected by others	.46
Factor XIV: 1.99 per cent	
Cold and unfeeling	.80
Spoils people with kindness	.48
Factor XV: 2.04 per cent	
Can complain if necessary	.80
Factor XVI: 2.52 per cent	
Obeys too willingly	.70
Factor XVII: 2.37 per cent	
Able to doubt others	.68
Can be obedient	.67
Factor XVIII: 2.22 per cent	
Self-seeking	.62
Proud and self-satisfied	.52
Outspoken	.49

significance at the .05 level. Only three main effects were significant at the .05 level and will not be considered. In all, 15 of 54 *F* ratios reached significance at the .01 level or beyond. (See Table 1.)

TABLE 1
F RATIOS FOR ANALYSES OF VARIANCE OF FACTOR SCORES[a]

	Source		
Factor	Instruction (I)	Occasion (O)	I × O
I	18.92***		1.31
II	7.04***	2.37	
III	2.61*		
IV	10.34***		1.31
V	3.64**		1.13
VI	6.08***	6.99**	
VII	4.19***	1.75	
VIII	1.26	2.91*	
IX	20.73***	7.51***	
X	5.08***		1.31
XI			
XII	8.85***		
XIII	3.40**		1.49
XIV	3.59**	1.62	1.18
XV	1.78		1.34
XVI	4.71***		1.46
XVII	1.31		
XVIII	2.27*	4.11**	

* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] *F*s less than unity are not tabulated.

Mean factor scores for the instructional sets are reported in Figure 1. Results are reported for only the 12 factors with which differences significant at the .01 level were found. Because there were no significant set by occasion interactions, each mean was computed from the factor scores of 17 *S*s on five occasions (a total of 85 scores): i.e., the main effect of instructional set is reported in Figure 1.

Mean factor scores for each occasion are presented in Figure 2. Again results are reported for only factors with which differences significant at the .01 level were found. Means were computed from factor scores of 17 *S*s with eight instructional sets (a total of 136 scores): i.e., the main effect of occasion is reported in Figure 2.

The items, with their factor loadings, reported above define the factors, of course. However, to simplify discussion of these factors, judgments were obtained from a sample of eight psychologists about the traits represented by each factor that yielded either a significant set or occasion main

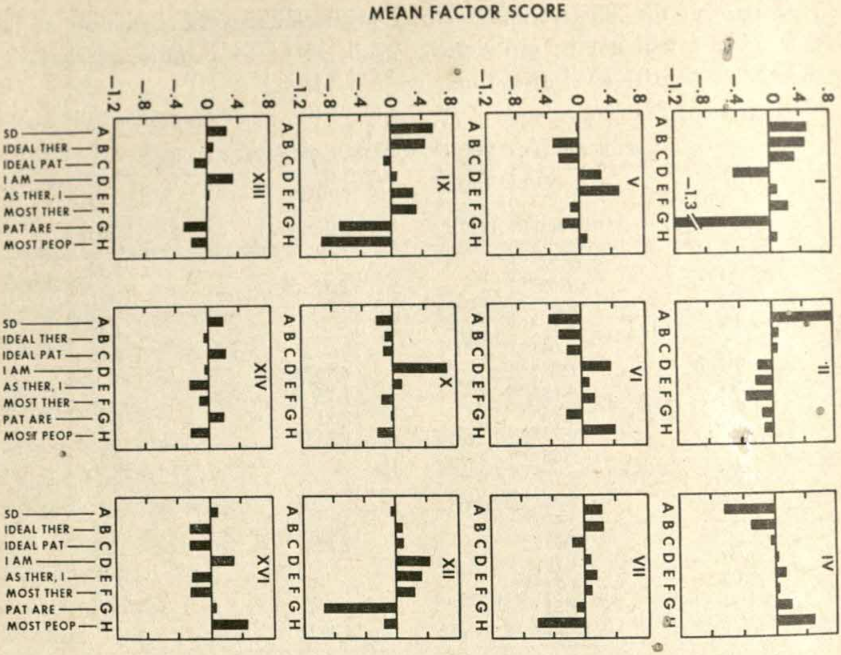

FIGURE 1
Mean Factor Scores as a Function of Instructional Set

effect. The writers then collated these judgments and agreed on the following factor names: Factor I: Self-Reliance, Factor II: Over-Friendliness, Factor III: Dependence, Factor V: Indulgence, Factor VI: Resentment of Supervision, Factor VII: Tough-Mindedness, Factor IX: Benevolent Leadership, Factor X: Sarcasm, Factor XII: Appropriate Firmness, Factor XIII: Admirable Competitiveness, Factor XIV: Rejection, Factor XVI: Excessive Obedience, and Factor XVIII: Self-Aggrandizement. In subsequent discussion, the factors will be referred to by these names.

D. Discussion

The absence of an interaction between occasion and instructional set for any of the 18 factors indicates, of course, that the relationship between various sets was stable over time: i.e., when a factor did change over time, all sets changed by equivalent amounts. In another experiment of similar design (1), significant interactions between occasions and sets were not observed either. With respect to Factors VI, IX, and XVIII (Figure 2), training apparently

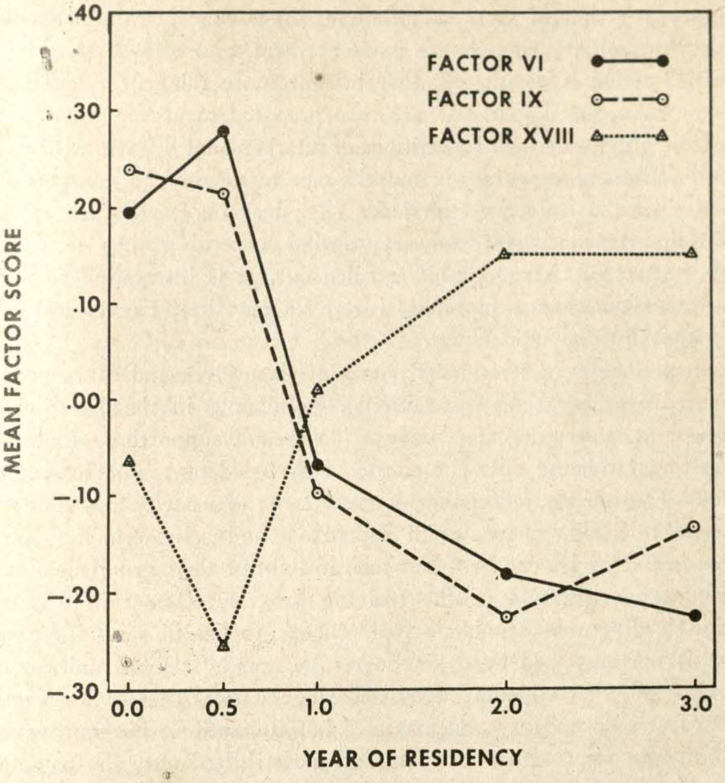

FIGURE 2
MEAN FACTOR SCORES AS A FUNCTION OF YEARS FROM THE BEGINNING OF THE PSYCHIATRIC RESIDENCY

had a pervasive influence on the residents' perceptions of interpersonal behavior instead of selectively modifying specific role conceptions such as "I am" or "Patients are." For example, the residents seemed to perceive less resentment of supervision (Factor VI) not only in themselves but also in patients, most people, the ideal therapist, etc. Furthermore, the highly significant set difference (Figure 1) indicates that this pervasive change is not due to a failure to discriminate between the sets.

The major part of the changes reported in Figure 2 occurred between the end of the first six months and the beginning of the third year. For the factors actually undergoing significant change, the first six months are a period of relative stability preceding the changes that ensue. This finding is consistent with our impression of residents during the early months of their training. It is a time of rapid assimilation of new concepts and attitudes (7), in which unfamiliar demands are made of them, and of self-appraisal. Once this initial period is weathered, they become more relaxed and confident in their new roles, and apparently are then prepared to accept modifications in some of the previously held conceptions of interpersonal behavior. The changes observed following the first six months can reasonably be considered to be a function of the residency experience (i.e., learning), although it was not possible simultaneously to study appropriate control groups so that other possible sources of change could be ruled out. It is interesting to note that the forms of the change functions (after six months), like typical learning curves, appear to be exponential.

More specifically, the residents' views of themselves and other persons, as represented by the various instructional sets, change in the direction of less resentment of supervision, less self-sacrificing and supporting of others, and more self-aggrandizing over the course of the residency. The former change (Factor VI) probably reflects an increased sense of security and confidence in their new role. The decreases in supportive and self-sacrificing leadership revealed in Factor IX can be understood in light of their experiences as therapists and the orientation of the training program. The initial view of a physician leading the psychiatric patient back to health with direction and support is not sustained by dramatic results and by the attitudes of supervisors. Instead, the residents learn from these sources that such an approach tends to foster dependency and extended hospitalization. The burden and responsibility for improvement is more and more shifted onto the patient, with the psychiatrist playing more of a helping than a leading role.

A few of the factors not showing significant change are noteworthy for that reason. One of these is self-reliance or confidence (Factor I). It might be reasonably expected that self-reliance would increase over training, as indeed other observations suggest it does. Whether the R-ICL and/or method of analysis is insensitive to such change even though it occurs, or whether the judgment of supervisors is largely biased by what they want to believe is indeterminate from the present data. The problem at least illustrates the necessity for further work on the relation of the residents' verbal reports to other behaviors.

As noted in the introductory comments, no comparable research on personality changes in psychiatric residents has been reported. Holt's (7) qualitative observations on such changes differed in two important respects from those reported above: (*a*) intrapersonal variables were emphasized in contrast with interpersonal variables in the present study, and (*b*) broad concepts (such as ego-identity) were used in contrast with the more specific concepts as defined by factors in the recent research. Since the variables observed were essentially different, the results of the two studies are not comparable.

Figure 1 presents data on how residents see themselves (I am), in general, and, more specifically, in their role as a therapist (As a therapist I am). In contrast with their views of what is desirable, they see themselves as lacking in self-reliance (I) and as being too indulgent (V) and too sarcastic (X). In comparison with themselves, patients are perceived as being much more lacking in self-reliance (I) and leadership (IX), and as being uncooperative (XIII), excessively obedient (XVI), and unable to assert themselves (XII). There are several dimensions on which the residents clearly distinguish between the way they see themselves (I am) and their role as a therapist (As a therapist I am). As therapists they consider themselves to be somewhat more self-reliant (I) and indulgent (V), but much less sarcastic (X), competitive (XVIII), and excessively obedient (XVI). In summary, residents report that they are lacking in some traits that constitute an ideal therapist, but, with the major exception of Factor V (Indulgence), they perform more like the ideal therapist in their role as a therapist than they do otherwise.

Although the results did not reveal huge changes in interpersonal perceptions during psychiatric training (only three of 18 main effects of occasion were significant at the .01 level or beyond), they were encouraging in relation to a previous analysis of scores on pre-established scales which did not yield significant differences. Most of the factors here obtained are found also in a similar analysis of data obtained with the same sample of 64 items but with different *S*s (psychiatric patients instead of residents) and with a sample of instructional sets that differed considerably (1). Thus, the flexibility gained by factor analyzing item responses does not seem necessarily to demand the price of low generalizability. In view of these results, it is concluded that Horst's (8) factor analytic approach to the problem of assessing changes over time is sufficiently promising empirically to justify further application and evaluation.

E. SUMMARY

The Revised Interpersonal Check List was administered to 17 psychiatric residents on five occasions (zero, five-tenths, one, two, and three years from the beginning of the residency) with eight instructional sets (I am, Most people are, etc.). The results were first factor analyzed by item in accordance with a model developed by Horst (8), then factor scores were computed, and finally the effects of occasions and sets were evaluated by analysis of variance.

Three of 18 main effects of occasion were significant ($p < .01$) and suggested that the residents' perceptions of themselves and others change in the direction of less resentment of supervision, less self-sacrificing and supporting of others, and more self-aggrandizing over the course of residency. Twelve of 18 main effects of instructional set were significant ($p < .01$) and suggested that the residents see themselves as lacking in self-reliance and being too indulgent in contrast with what they consider desirable. It was concluded that Horst's new factor analytic approach to assessing change was promising, particularly in view of the failure of different analyses of these data to yield significant findings.

REFERENCES

1. BOE, E. E., GOCKA, E. F., & KOGAN, W. S. The effect of group psychotherapy on interpersonal perceptions of psychiatric patients. *Multivariate Behav. Res.*, (in press).
2. EDWARDS, A. L. The Social Desirability Variable in Personality Assessment and Research. New York: Dryden, 1957.
3. FOA, U. G. Convergences in the analysis of the structure of interpersonal behavior. *Psychol. Rev.*, 1961, **68**, 341-353.
4. GOCKA, E. F. The pseudo variate method of analyzing nested design data. Report RS-027-65, Western Research Support Center, VA Hospital, Sepulveda, California, 1965.
5. HARMAN, H. H. Modern Factor Analysis. Chicago: Univ. Chicago Press, 1960.
6. HARRIS, C. W., *Ed.* Problems in Measuring Change. Madison: Univ. Wisconsin Press, 1963.
7. HOLT, R. R. Personality growth in psychiatric residents. *Arch. Neurol. Psychiat.*, 1959, **81**, 203-215.
8. HORST, P. Multivariate models for evaluating change. In C. W. Harris (Ed.), *Problems in Measuring Change.* Madison: Univ. Wisconsin Press, 1963. Pp. 104-121.
9. KOGAN, W. S., & FORDYCE, W. E. The control for social desirability: A comparison of three different Q sorts and a check list, all composed of the same items. *J. Consult. Psychol.*, 1962, **26**, 26-30.
10. KRIEGER, M. H., & KOGAN, W. S. A study of group processes in the small therapeutic group. *Internat. J. Group Psychother.*, 1964, **14**, 178-188.
11. LAFORGE, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal checklist. *J. Personal.*, 1955, **24**, 95-112.
12. LINDQUIST, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin, 1953.

Veterans Administration Hospital
4435 Beacon Avenue So.
Seattle, Washington 98108

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 241-248.

DISTRIBUTION OF ANAGRAM SOLUTION TIMES*[1]

Ophthalmology Department, New York University Medical Center

IRA T. KAPLAN, THOMAS CARVELLAS, AND BARBARA J. BREININ

A. INTRODUCTION

Previous studies by Mayzner and Tresselt (3) and Hunter (2) have suggested that the distribution of solution times for anagram problems has a characteristic shape. The cumulative distribution of solution times is a negatively accelerated increasing function of time: i.e., the longer solution times are increasingly rare. In other words, the longer *S* goes without finding a solution, the more additional time he will probably take to find one.

Mayzner and Tresselt proposed that the distribution of solution times reflects the rate at which *S* generates attempts at solution. Thus, when an anagram is first presented it suggests a variety of letter combinations and words, but as *S* continues to work he becomes progressively slower at thinking of new potential solutions. In support of their hypothesis Mayzner and Tresselt cited the word-association experiments of Bousfield and Sedgewick (1), which showed that associations are produced rapidly at first but become progressively slower with the passage of time.

If the distribution of anagram solution times has literally the same shape as the distribution of associative response times, then the same mathematical formula should describe both kinds of data. In the present study we measured the time that *S*s required to solve various five-letter anagrams. Our purpose was first to find an equation that would describe our data, then to discover whether that equation would fit the anagram solution times reported by Mayzner and Tresselt and by Hunter, and whether it would also describe the production of verbal associates reported by Bousfield and Sedgewick.

B. METHOD

1. *Subjects*

The *S*s were nine undergraduate males from an introductory psychology course and one female graduate student in psychology.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 2, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by Public Health Service Research Grant MH 08146 from the National Institute of Mental Health.

2. *Stimulus Materials and Apparatus*

Each *S* was presented with 40 five-letter anagrams. The anagrams, typed in capital letters on slides, were projected individually onto a screen by means of a Kodak Carousel projector. The anagrams were formed by randomly rearranging the letters in each of 40 common words, taken largely from the Rees and Israel (4) list of anagrams with single solutions. Solution times were measured by an electric timer, which was started by a photocell as soon as the anagram appeared on the screen, and was stopped when *S* pressed a telegraph key on arriving at a solution.

3. *Procedure*

The *S*s were told that each group of letters presented on the screen could be rearranged to form a common five-letter word. They were instructed to begin solving each anagram as soon as it appeared, to press the key immediately upon finding a solution, and then to announce their solution to *E*. The solution time was recorded, and the next anagram was presented. If no solution was found within 100 seconds, the trial was terminated by presentation of the next anagram. Every *S* received the same set of 40 anagrams during the session, but the order of presentation was randomized for each *S*. All *S*s were tested individually.

C. RESULTS

The cumulative distribution of solution times for all 10 *S*s to 40 anagrams is shown in Figure 1. The points are based on data from 389 trials. Eleven trials were lost because seven times *S* mistakenly pressed the key when he did not have a solution, and four times the solution offered was different from that found by the other *S*s.

The broken line in Figure 1 is an exponential curve that Bousfield and Sedgewick (1) fitted to their data on word associations. Its equation is

$$n = c\,(1 - e^{-mt}), \qquad [1]$$

where n = the cumulative frequency of responses, c = an empirical constant that determines the limit approached by the curve, e = the base of the natural logarithms, m = an empirical constant that determines the rate at which the limit is approached, and t = time.

The curve was fitted so that it passed through the 50- and 100-second points according to the method used by Bousfield and Sedgewick, which estimates $c = 84.9$ and $m = .050$. This method achieves a good fit for the asymptotic portion of the curve at the expense of a marked discrepancy for

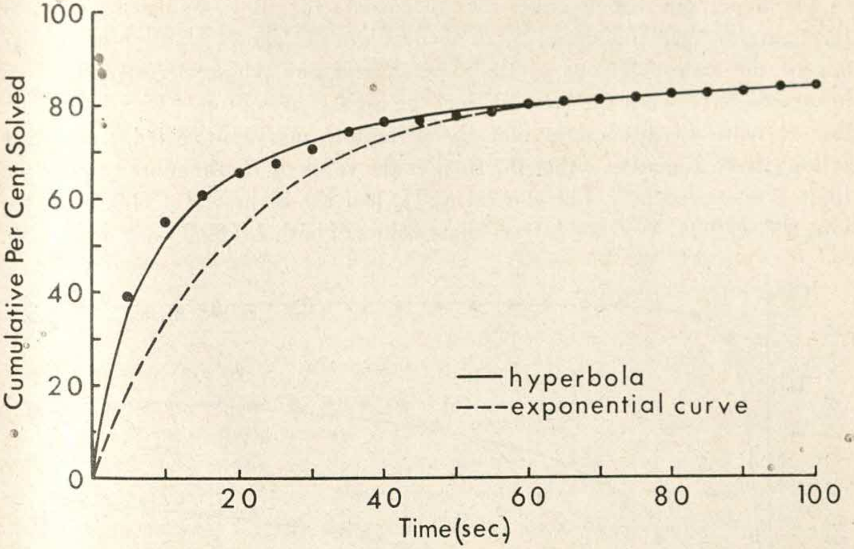

FIGURE 1

CUMULATIVE DISTRIBUTION OF ANAGRAM SOLUTION TIMES

The points represent pooled data from 10 *S*s of the present experiment. The solid line is an hyperbola with constants $L = 90.7$ and $R = 7.5$. The broken line is an exponential curve with constants $c = 84.9$ and $m = .05$.

the shorter solution times. The basic reason for this discrepancy, however, is that the data do not follow an exponential curve. A semilog plot of the data, according to the conventional method for reducing an exponential curve to a straight line, results in a curved line rather than a straight one.

The solid line that does fit the data in Figure 1 is an hyperbola with the equation,

$$n = \frac{Lt}{t + R}, \qquad [2]$$

where $n =$ the cumulative frequency of responses, $L =$ an empirical constant that determines the limit approached by the curve, $R =$ an empirical constant that determines the rate at which the limit is approached, and $t =$ time. This equation was used by Thurstone (5) as a learning curve. Its constants L and R are estimated graphically by plotting t/n against t, which reduces the hyperbola to a straight line, as it did for the present data. The constant $L = 90.7$ is the reciprocal of the slope, and the constant $R = 7.5$ is the absolute value of the X intercept of that straight line. When these values were substituted in Equation 2, they produced the hyperbola of Figure 1.

The hyperbola that describes the pooled data for all 10 *S*s can be fitted to the data for each individual *S*, as well. Figure 2 shows solution times for one of the fastest *S*s, one of the slowest, and one whose performance was intermediate between the two. SE had the highest asymptote ($L = 100$). His low *R* value of .6 signifies that the asymptote was approached rapidly. It follows from Equation 2 that the smaller the value of *R*, the more rapidly the limit *L* is approached. The slowest *S*, JI, had the highest *R* (34.0) and the lowest *L* (79.0). WZ had intermediate values of both *L* (89.3) and *R* (11.0).

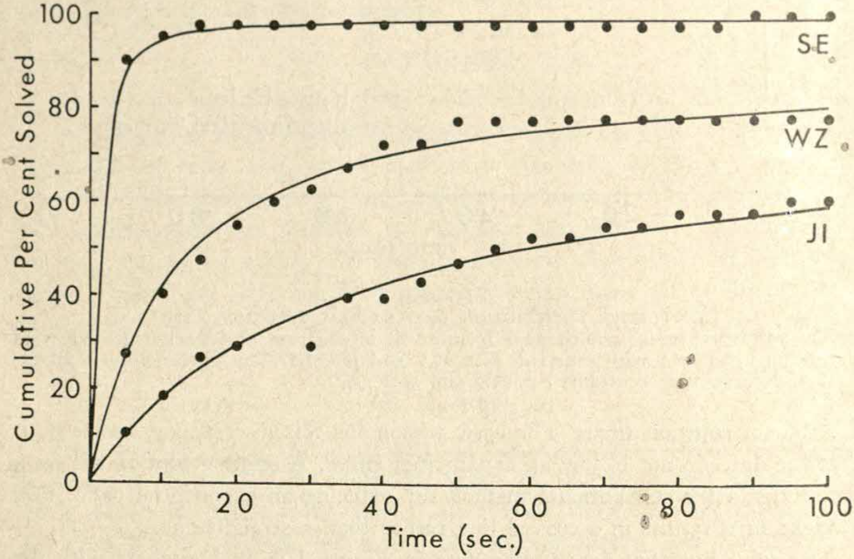

FIGURE 2

Cumulative Distribution of Anagram Solution Times for Three *S*s of the Present Experiment

The constants for SE are $L = 100.0$, $R = .60$; for WZ, $L = 89.3$, $R = 11.0$; for JI, $L = 79.0$, $R = 34.0$.

Notice that there is a negative correlation between *L* and *R*. A high value of *L* is associated with a low value of *R*. This relation holds for all 10 *S*s, whose constants are summarized in Table 1. The product moment correlation coefficient (*r*) relating *R* and *L* is equal to —.71.

D. Discussion

Now let us consider whether the hyperbola that fits the present results describes the anagram solution times gathered in earlier studies. Then we

TABLE 1
CONSTANTS OF HYPERBOLAS FITTED TO THE DATA OF INDIVIDUAL *S*s

Subject	*L* (% solved)	*R* (sec.)
SE	100.0	0.6
DL	100.0	1.0
BL	100.0	4.0
MS	100.0	11.0
GN	97.1	11.0
JS	90.7	8.0
WZ	89.3	11.0
JA	86.6	8.0
JI	79.0	34.0
JM	70.5	17.1

will see whether it fits the association data also. Figure 3 shows clearly that the hyperbola describes the anagram data far better than the exponential curve. The three cumulative distributions in the figure are from the Mayzner and Tresselt (3) study on the effects of letter order and word frequency on solution time. The label Easy Order indicates that the letters of the anagrams were only slightly rearranged from their positions in the solution words, while the term Hard Order means that they were thoroughly rearranged. The label VF means that the words occurred very frequently in the Thorndike-Lorge word count, and VI stands for very infrequently. Note that these authors collected solution times up to 240 seconds, *versus* 100 seconds in the present study, and the hyperbola continues to fit.

The same negative correlation between L and R that was found for individual differences in the present study seems to hold for differences in anagram difficulty as well. The easiest anagrams have the highest limit, which is approached at the fastest rate. Remember that a low value of R causes the limit to be approached rapidly. For the three curves of Figure 3, in order of increasing difficulty, the constants are $L = 100.0$, $R = 5.0$; $L = 93.5$, $R = 25.7$; $L = 72.1$, $R = 51.9$. Hunter's data (2), which extends only to 30 seconds, fits in with the same trend. His results follow the form of a hyperbola, whose constants of $L = 96.8$ and $R = 9.3$ place it between the upper two curves of Figure 3.

Bousfield and Sedgewick (1) used an exponential equation to describe the production of verbal associates. This equation fitted some of their data, but in most cases there was an initial divergence between the data and the fitted curve, such that the early part of the curve fell below the data points. In those cases where the exponential curve did fit the data, we found that the hyperbola fitted just as well. Figure 4 illustrates the more common result: initial divergence between the data and the exponential curve. Here the

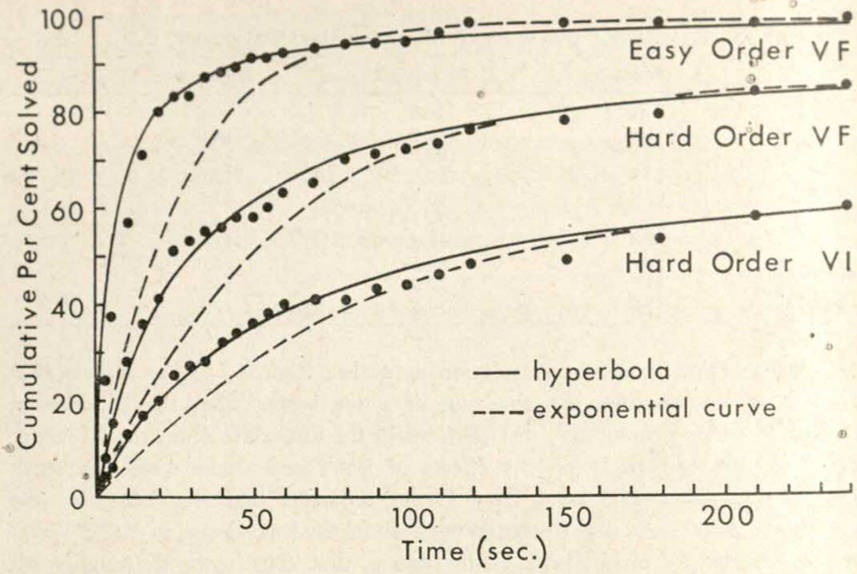

FIGURE 3
CUMULATIVE DISTRIBUTIONS OF SOLUTION TIMES FOR THREE GROUPS OF ANAGRAMS FROM DATA OF MAYZNER AND TRESSELT[2]

The constants of the hyperbolas for Easy Order VF are $L = 100.0$, $R = 5.0$; for Hard Order VF, $L = 93.5$, $R = 25.7$; for Hard Order VI, $L = 72.1$, $R = 51.9$. The constants for the corresponding exponential curves are $c = 99.0$, $m = .0382$; $c = 86.2$, $m = .0178$; and $c = 64.0$, $m = .0116$.

hyperbola clearly gives a better fit. Thus, we can conclude that the same equation describes both the distribution of anagram solution times and the production of verbal associates. This equation, however, is not the exponential curve that Bousfield and Sedgewick originally applied to their data, but it is the equation for an hyperbola.

Although the hyperbola describes both types of data, two differences should be noted between the results for anagrams and those for associates. One is that, for the particular problems used in these studies, the limit of associative responses was approached much more slowly than the limit of anagram solutions. The estimated values of L and R for all 14 conditions of the association study are summarized in Table 2. Notice that R is expressed in minutes in Table 2 and in seconds in Table 1. When converted into minutes the range

---

[2] This figure is based on data presented in Figure 1 of Mayzner and Tresselt (3), and is reproduced by permission of Dr. Mayzner and of the American Psychological Association.

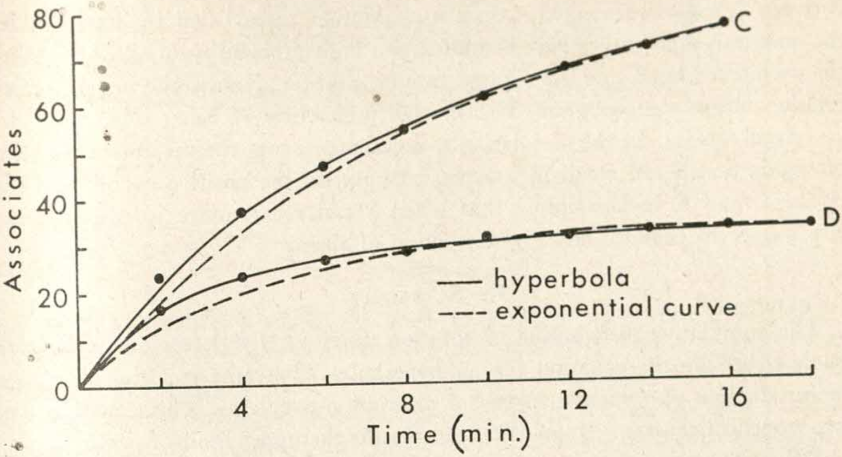

FIGURE 4
CUMULATIVE DISTRIBUTIONS OF ASSOCIATIVE RESPONSES FROM DATA OF BOUSFIELD AND SEDGEWICK

Plot C was obtained under instructions to name U. S. cities, plot D under instructions to list makes of automobiles. The exponential curves that Bousfield and Sedgewick fitted to their data have constants $c = 92$, $m = .110$ for C; and $c = 34.3$, $m = .219$ for D. The constants of the hyperbolas are $L = 118.6$, $R = 8.9$ for C; and $L = 37.8$, $R = 2.4$ for D.

of R for anagrams goes only from .01 to 1.3 minutes, while the range of R for associates extends from 1.5 all the way to 84.1 min.

The second difference is that in contrast to the high negative correlation between L and R for anagrams, there is a high positive correlation ($r = .91$)

TABLE 2
CONSTANTS OF HYPERBOLAS FITTED TO THE DATA OF BOUSFIELD AND SEDGEWICK

Instructions	L (number of responses)	R (min.)
5-letter words	470.6	84.1
Fellow college students	195.1	18.3
U. S. cities	118.6	8.9
Pleasant objects, activities, and situations	116.7	15.9
Quadruped mammals	70.2	4.9
Unpleasant objects, activities, and situations	70.0	12.4
Words from QNDTERH	53.3	10.3
Animals—names with two or more syllables	50.5	8.1
Words containing M, T, D	48.2	31.8
European cities	47.8	5.8
Friends—surnames beginning with S	45.6	10.5
Birds	39.5	4.6
Makes of cars	37.8	2.4
Makes of cars, with preliminary learning	35.3	1.5

between L and R among the various association tasks listed in Table 2. In the anagram problems, a high L indicates a high probability of ultimate solution, while a low R signifies a rapid rate of solution. Hence, the negative correlation means that anagrams with a high probability of being solved tend to be solved rapidly. In the association tasks, L represents the maximum number of words that S can recall in a given category, so the positive correlation between L and R means simply that when S can recall more words, it takes him longer to produce any given fraction of them.

E. Summary

The cumulative distribution of solution times for five-letter anagrams can be described by the equation for an hyperbola. The same equation describes the production of responses in verbal association problems. This equation has two empirical constants, L and R. L represents the upper limit of performance, and R is inversely related to the rate at which this limit is approached. Individual differences in anagram-solving ability were described in terms of variation in L and R. There was a negative correlation ($r = -.71$) between these constants, so that the higher S's limit, the more rapidly he approached it. A similar relation held for differences in anagram difficulty: the easier anagrams had a higher limit, which was approached more rapidly. The opposite relation ($r = .91$) held among different association problems: in a category where S could produce more responses, he approached this limit relatively more slowly.

References

1. Bousfield, W. A., & Sedgewick, C. H. W. An analysis of sequences of restricted associative responses. *J. Gen. Psychol.*, 1944, **30**, 149-165.
2. Hunter, I. M. L. The solving of five-letter anagram problems. *Brit. J. Psychol.*, 1959, **50**, 193-206.
3. Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *J. Exper. Psychol.*, 1958, **56**, 376-379.
4. Rees, H. J., & Israel, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monog.*, 1935, **46**, No. 6 (Whole No. 210).
5. Thurstone, L. L. The learning curve equation. *Psychol. Monog.*, 1919, **26**, No. 114.

Department of Ophthalmology
New York University Medical Center
New York University School of Medicine
550 First Avenue, New York, N. Y. 10016

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 249-254.

A NOTE ON TWO BASIC FORMS OF CONCEPTS AND CONCEPT LEARNING*

Research and Development Center for Learning and Re-education, University of Wisconsin

GARY A. DAVIS

A. INTRODUCTION

Concept learning is probably the most important of all instances of learned human behavior. Indeed, concept learning has been of theoretical and practical interest to psychologists and educators for decades and to philosophers for centuries. There has been, however, a marked failure, or at least a gross obscurity, in clarifying some very basic questions: What is concept learning? What is a concept? How are concepts learned?

The purpose of the present note is to offer a suggestion concerning what appears to be two basic types of "concept learning" and "concepts." It will be seen that the two forms of concept learning differ in respect to such important dimensions as (*a*) the nature and definition of "a concept," (*b*) basic principles of acquisition, (*c*) the role of reinforcement, (*d*) the nature of the stimulus, (*e*) the nature of the response, (*f*) what is learned, and (*g*) relevant experimental literature.

B. CONCEPT LEARNING I: CATEGORIZING BEHAVIOR

The first type of concept learning consists of learning the attributes (or characteristics or cues) which define a category. Once the defining attributes have been acquired, new instances of the concept or category can be appropriately identified as such. Since the attributes are common to all instances of a given concept, these cues are considered to be *abstract* rather than concrete. Thus several researchers have referred to this type of concept learning as a process of abstraction or abstraction learning (5, p. 75). It is obviously this form of concept learning as categorizing behavior whose laboratory analogue is the card-sorting experiment in which *S* learns to classify geometric-patterned stimuli into usually two categories on the basis of some predetermined combination of stimulus characteristics: e.g., *red triangles* might belong in one category, and all other patterns in the other category.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 4, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

Concept learning as categorizing behavior is schematized in the left portion of Figure 1. Symbolized here is simply the notion that, once the concept is learned, different stimulus objects or situations which possess the common defining attributes will be categorized together.

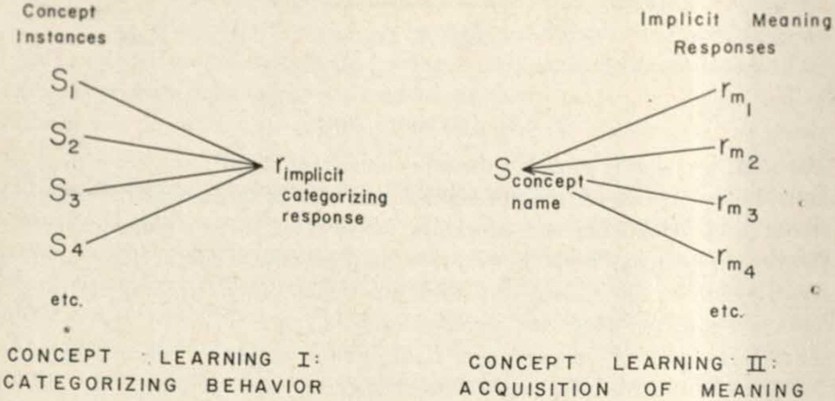
Concept Instances
S_1
S_2
S_3
S_4
etc.
$r_{implicit}$ categorizing response
CONCEPT LEARNING I: CATEGORIZING BEHAVIOR
Implicit Meaning Responses
$S_{concept}$ name
r_{m_1}
r_{m_2}
r_{m_3}
r_{m_4}
etc.
CONCEPT LEARNING II: ACQUISITION OF MEANING

FIGURE 1
DIAGRAM OF CONCEPT LEARNING AS CATEGORIZING BEHAVIOR (CONCEPT LEARNING I) *vs.* CONCEPT LEARNING AS ACQUISITION OF MEANING (CONCEPT LEARNING II)

"The concept" in this form of concept learning is the category itself and defining abstract properties, from which the category is logically inseparable.

Whether one speaks of forming concepts in childhood, learning concepts in the classroom, or identifying "concepts" in the card-sorting experiments, in all cases correct categorizing responses to those stimulus instances possessing the defining attributes are selectively reinforced, and incorrect categorizing responses are not reinforced. Thus the basic learning principles which underlie this form of concept learning as categorizing would be those of *instrumental conditioning*, with reinforcement serving primarily a *feedback* function rather than a motivational function. Accordingly, the conditions for learning the attributes which define a category would be the contiguous presentation of concept instances with feedback indicating membership or nonmembership in the category.

In the card-sorting experiments, using either the sequential (1) or simultaneous (2) stimulus presentation method, *S*s always make overt categorizing responses prior to the reinforcing feedback. It is likely, however, that the concept could be learned without any overt response. Indeed, in everyday classroom teaching concept instances are presented in conjunction with appropriate feedback indicating membership or nonmembership in the category, and *S*s learn without overtly responding to each instance. In general, congruent with the intuitive notion of conceptual behavior as implicit, the categorizing response of importance would be an implicit categorizing response rather than any overt response. The implicit categorizing response can be assumed to control somehow any overt motor (including verbal) categorizing responses which may occur.

Consideration of classroom learning suggests another (often supplementary) method of teaching concepts which would be entirely inappropriate for card-sorting situations—the presentation of a verbal definition of the concept. The inappropriateness of presenting college *S*s with a verbal definition of the "concept" of *red triangles,* and then testing their grasp of this concept by asking them to recognize new instances, exemplifies the lack of complete generality of current laboratory concept identification experiments to real-world and classroom concept formation.

C. Concept Learning II: Acquisition of Meaning

Another type of concept learning, which seems to be at least partially independent of categorizing behavior, is the acquisition of meaning responses to a stimulus. The stimulus in this case is usually a word. For example, the stimulus words or concept name *Abraham Lincoln* when presented to normal English-speaking adults will elicit an elaborate network of implicit responses which take the form of verbal and nonverbal (including affective) associations. This distribution of responses is the associative meaning of the stimulus word and can be interpreted as the associative concept named by the word (3).

"The concept" in this form of concept learning is the stimulus word (concept name) in conjunction with the network of implicit meaning associations elicited by that word, and one is said to have formed the concept when the word comes to elicit the appropriate associative responses.

Concerning the basic conditions for the acquisition of meaning responses to a concept name, the contiguous occurrence of the word with the sensory stimulation to be associated with that word, given attention, should be the necessary and sufficient conditions for learning. For example, the contiguous occurrence of the concept name *Abraham Lincoln* with pictures of Lincoln, words or pictures related to the concepts of slavery, the Civil War, the president of the United States, positively (or negatively) toned words, etc., would lead to the acquisition of the concept of *Abraham Lincoln.* Thus principles of *classical conditioning* would seem to describe concept learning as the acquisition of meaning responses to a concept name.

Concerning the most relevant experimental literature, papers dealing with verbal associations, classical conditioning of word meaning [e.g., (6)], and the semantic differential measurement of word meaning would seem appropriate. Reinforcement (in either a feedback or motivational form) in these research areas, and in this second form of concept learning as the acquisition of meaning responses, is mostly irrelevant to the basic condition of contiguous occurrence of the concept name and the to-be-learned units.

TABLE 1
A SUMMARY OF DISTINCTIONS BETWEEN TWO TYPES OF CONCEPT LEARNING

Variable	Concept Learning I: Categorizing	Concept Learning II: Acquisition of meaning
What is "The Concept"?	A category and the defining cues	A word (concept name) and the associated verbal and nonverbal meaning responses
Basic principles of acquisition	Instrumental conditioning	Classical conditioning
Conditions for learning	Contiguous presentation of concept instances with appropriate feedback and/or verbal definition	Contiguous presentation of concept name with other verbal and nonverbal stimuli
Role of reinforcement	Important in feedback function	Irrelevant
Nature of the stimulus	Abstract cues	Concept name (word)
Nature of the response	Implicit categorizing response	Implicit network of verbal and nonverbal associations
What is learned?	Abstract cues which define the concept	Implicit verbal and nonverbal responses are conditioned to the concept name
Experimental literature which may be relevant	Concept identification experiments of the card-sorting type	Experiments dealing with verbal associations, verbal clustering, semantic differential measurement of meaning, and classical conditioning of meaning

The major distinctions between concept learning as categorizing behavior *vs.* concept learning as the acquisition of meaning are summarized in Table 1.

D. DISCUSSION

It might appear that the two types of concept learning, as shown in Figure 1, are simply the reciprocals of one another. In other words, with concept learning as categorizing behavior, specific concept instances with their defining attributes will elicit the categorizing response *r*, part of which is the concept name. Conversely, the concept name when presented as a stimulus would elicit as associations exactly those defining attributes and concept instances. Admittedly, to the extent that the meaning responses or associations elicited by a concept name are nothing but criterial attributes and instances of the concept, then the present proposal of a distinction between concept learning as categorizing *vs.* acquiring meaning would certainly be unjustified.

A closer look via some examples of what typically are considered to be "concepts" would suggest, however, that the two types of concept learning are not completely mirror images or reciprocals of each other, particularly with more complex concepts. For example, with relatively simple concepts or abstractions, such as *white, round,* or even *tangent, animal,* and *flower, S* would learn to categorize concept instances on the basis of criterial attributes and, once the concept is learned, the concept name would then elicit as associations the same attributes and instances of that concept. At the other extreme, the acquisition of such complex concepts as *industrial revolution, Abraham Lincoln,* and *education* is more a matter of acquiring meaningful associations to the concept name than of learning to categorize concept instances on the basis of common cues: i.e., the reciprocal relationship described above would not be perfect.

In any event, if these two forms of concept learning can be justified, are there perhaps more? While the present scheme may not exhaust all possible forms of concept learning, it would seem to encompass most. For example, the learning of *relational concepts* (e.g., *larger than, part of, descended from, equal to,* etc.) could be considered as an instance of categorizing behavior in which all members of the concept possess the common defining attribute: i.e., the particular relationship. Also, such an obscure type of concept learning as the learning of *theoretical concepts* (e.g., the concepts of *evolution, Gestalt psychology, theory of action,* etc.) would be an example of concept learning as the acquisition of meaning in which the concept, if one has formed the concept, will elicit a network of implicit verbal and nonverbal responses.

In sum, the present analysis of concept learning is intended as a simplifying scheme for distinguishing between two major usages of the term "concept learning." While this conception is by no means completely original or unique, and in fact may be implicit in some mediational models of human behavior [e.g., (4)], recognizing that concept learning may be used with reference to categorizing behavior *vs.* acquisition of meaning, which may differ along the dimensions summarized in Table 1, would seem to reduce much of the ambiguity of the term "concept learning."

E. SUMMARY

Concept learning as categorizing behavior is distinguished from concept learning as the acquisition of word meaning. These two types (and usages) of "concept learning" are shown to differ along several important dimensions such as the relevant principles and the conditions of learning. Learning to categorize concept instances is largely a matter of the selective reinforcement of correct categorizing responses to stimuli possessing the defining cues and nonreinforcement of incorrect categorizing responses (instrumental conditioning). On the other hand, the acquisition of meaning responses to a concept (word) occurs as a result of contiguous presentation of the word with the items to be associated with that word (classical conditioning).

REFERENCES

1. BOURNE, L. E., JR., & BUNDERSON, C. V. Effects of delay of informative feedback and length of postfeedback interval on concept learning. *J. Exper. Psychol.*, 1963, **65**, 1-5.
2. BRUNER, V. S., GOODNOW, J. J., & AUSTIN, G. A. A Study of Thinking. New York: Wiley, 1956.
3. DEESE, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161-174.
4. STAATS, A. W. Verbal habit-families, concepts, and the operant conditioning of word classes. *Psychol. Rev.*, 1961, **68**, 190-204.
5. STAATS, A. W., & STAATS, C. K. Complex Human Behavior. New York: Holt, 1963.
6. STAATS, A. W., STAATS, C. K., & HEARD, W. G. Denotative meaning established by classical conditioning. *J. Exper. Psychol.*, 1961, **61**, 300-303.

Research and Development Center for
Learning and Re-Education
The University of Wisconsin
936 West Johnson Street
Madison, Wisconsin 53706

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 255-266.

A STUDY OF CERTAIN VISUAL EFFECTS OCCASIONED BY FACTORS OF SO-CALLED GLARE*[1]

Department of Psychology, Michigan State University

RICHARD F. HAINES[2] AND S. HOWARD BARTLEY

A. INTRODUCTION

The present study stemmed from the fact that the photic flux entering the eye is not confined to forming the expected image of the visual field and the visual target that is fixated. The junctures of the various internal media of the eye disperse a certain amount of the flux, and on some occasions there is a sizeable amount of flux that passes through the walls of the eye. All of this flux is spoken of as stray radiation[3] or, more commonly, as "stray light." This stray radiation has unwanted effects and, in some cases, these have influenced sensory end results (vision) in unrecognized ways.

Since there is no strict correspondence between the distribution of radiation in the retinal image and the distribution of light and dark in the target, certain problems arise both in the laboratory and in everyday practice. With a visual field containing a luminous target, the result is that not only is the expected object seen, but also a dimly lighted area around it. This shows up most strikingly in studies using intermittent inputs (flicker experiments) in which the supposedly unilluminated field around the object, as well as the object, flickers.

When the visual field contains two luminous targets, the stray radiation associated with each induces certain effects upon the other. For example, the threshold of visibility of the target fixated may be measurably raised (3, 15). This is an effect that the observer may not be aware of; but, nevertheless, it influences his visual efficiency. This effect has been labelled glare (10, 11, 12, 13).

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 7, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This work was performed while the principal author was supported by an N.I.H. Predoctoral Research Fellowship (1F1 MH-16,907-O1A1), Michigan State University, East Lansing, Michigan.

[2] Now at N.A.S.A., Ames Research Center, Moffett Field, California.

[3] In this paper we shall make a verbal distinction between the stimulus and the response, so that the word *light* will not refer to two very different things. The stimulus will be called photic radiation (or just radiation or flux) and what is seen will be called light.

Various factors are often at work to reduce visual efficiency or to provide visual annoyances. In vision, one is generally fixating on some portion of the visual field. This portion is called the target. Not only are there these fixated portions; but other portions, that, for want of a better term, can be called targets, too. Thus, there are the fixated target and peripheral targets.

Glare is also said to occur when a peripheral target is seen as something dazzling or annoying. In this case, glare is an *experience or response,* not an input or stimulus. Sometimes, however, the word glare is used to label the *photic input* that causes the experience just called glare. This results in the kind of ambiguity already mentioned. There is still a third use of the term—namely, to label the result (*rise in threshold* for seeing the fixated target), which may be unaccompanied by the glare experience or, at least, which is not the glare experience itself. The fourth use of the word is to label the *photic energy* that produces this rise in visual threshold.

While the matters we have dealt with in the present study may be included in the general area of glare, all phenomena in this area are not strictly those of glare. The terminology commonly used is actually more indefinite, loose, and ambiguous than it should be to promote adequate communication. On this account, we are injecting at this point what we hope is some clarification in terminology.

In most laboratories, the issue is not generally so much whether the peripherally located photic source produces the dazzling effect (glare) as whether it raises threshold for the fixated portion, since generally the measurement made has to do with either the observer's detection of this portion (the absolute threshold for it) or with its observed brightness. Thus, the thing called glare (as an experience) is typically not measured at all.

We have adopted the policy of calling only the annoying or irritating experience "glare." The labelling of the remaining factors is as follows: (*a*) the photic source or input not being fixated directly is called the *peripheral source;* (*b*) when a photic source produces glare, it is acting as a *glare source* no matter where it is located in the visual field; and (*c*) when it elevates the visual threshold for seeing any other target or alters any other visual function, it is called an *interference source,* since the action is a form of interference and not glare. There are two processes (and results) and two photic sources involved. Each item is given its own label and need not be confused with other items.

In the present investigation, since the fixated target is the highly luminous one, it is in effect the glare source; and what is being measured is its effect on seeing a small moving peripheral source.

Since the moving target moves in and out from behind the glare source, its position when it disappears and reappears is an indicator of the effective size of the glare source.

The impetus for the present study lay in the desire to resolve some questions about the effect of intense photic sources of small visual angular subtense in otherwise unilluminated fields, such as small, bright objects in a moonless sky. Their apparent locations and effects on each other are not too well known (17).

While the study as designed is not primarily one on glare, the reason for discussing glare is based upon the similarities between studies on so-called glare and the present study. These studies must be referred to; and to do this without confusion, a systematic use of language had to be devised.

B. The Problem

The problem that we wished to study stems from several considerations:

1. That various phenomena attest to the fact that the distribution of radiant flux on the retina is not strictly in the form generally assumed or idealized. A visual target which does not appear to taper gradually into its visual surrounds, but has abrupt borders, is represented on the retina by a gradient of illumination, the nature of which has been studied by several workers (2, 3, 4, 7, 9).

2. This gradient of illumination results in seeing a "halo" of weak light around a luminous target on an unilluminated or low luminance field.

3. This illumination gradient may produce a veil of radiation on the portions of the retina containing the images of other targets and, in turn, the illumination gradients produced by them may cast a veil of radiation on the retinal image area for the fixated target.

4. Likewise, the existence of the stray illumination producing retinal illumination gradients is evidenced in *electroretinograms* in various ways. For example, it has been found that the *stray* flux from a highly luminous target of small visual angle is the prime (almost exclusive) source of the B-wave in the record (6, 14) and that only when the visual angular subtense of the target is increased does the image illumination rather that the stray illumination predominate in producing the record.

Our specific problem is: How will the distribution of photic flux from two targets—a *stationary fixated one* and a less intense *moving target* that moves toward the fixated target in the frontal plane and passes behind it—affect the observation of the moving target?

In the test situation just described, the fixated target (in this case the glare and interference source) is far more intense than the moving target. Hence, it is only the taper (stray illumination) produced by the fixated target that is at issue. It would be supposed that the gradient of stray illumination that this target produces would cause the moving target to disappear behind the fixated target before otherwise expected, and reappear later than otherwise expected. This is to say, the moving target would be expected to disappear behind the veil of stray illumination that, of course, surrounds the otherwise expected image of the target. Thus, the disappearance of the moving target as it moves toward the fixated target (glare source) would be as if it were disappearing behind a bigger target than the metric dimensions indicate. The moving target would reach the gradient-determined boundary sooner than it would reach the actual border of the target.

Another way of viewing this situation is that the luminance of the moving target and that portion of the visual field where the moving target becomes invisible are equal—the contrast ratio (c) at this location being zero.[4]

It is the empirical measurement of this sort of discrepancy that has concerned us in this study. In order to provide more information concerning the effects of various shaped glare sources three different shapes were used instead of only one.

C. METHOD AND PROCEDURE

The experimental apparatus consisted of the glare source mentioned in the preceding section. The luminous source was a 1000 watt (3250°K) projection lamp housed in a box whose only opening for present purposes was a 1/2 inch diameter hole. The image of this hole was the source for filling the front surface of a tiny mirror (.155 square inch in area) which the observer fixated at 19.5 inches directly in front of him. Each of the three different shaped mirrors had the same frontal area in order to equate total photic flux to the eye, each subtended wholly foveal visual angles, and each had 15° back bevels on all sides. The angle of incidence onto and reflection from the mirror was 13° as indicated by α in Figure 1. The beam from the aperture of the lamp house just mentioned was collimated and filtered by neutral density Wratten filters. The maximum luminosity of this source was 4248.7 ft L. and the .4 log luminosity was 1699.4 ft L.

At the eye, there was a 3-1/2 mm diameter artificial pupil to occlude various extraneous sources of stray radiation. In front of the artificial pupil was an electric shutter to provide for temporally controlled exposures.

---

[4] The contrast ratio (c) is calculated by the formula $c = B_t - B_b/B_b$, where B_t = moving target luminance and B_b = stray illumination produced by the target luminance.

It was not assumed that a completely steady state of retinal adaptation would exist during any series of observations on account of the relatively brief periods of viewing and nonviewing.

In Experiment One, viewing time was 33 per cent of the fixed total trial

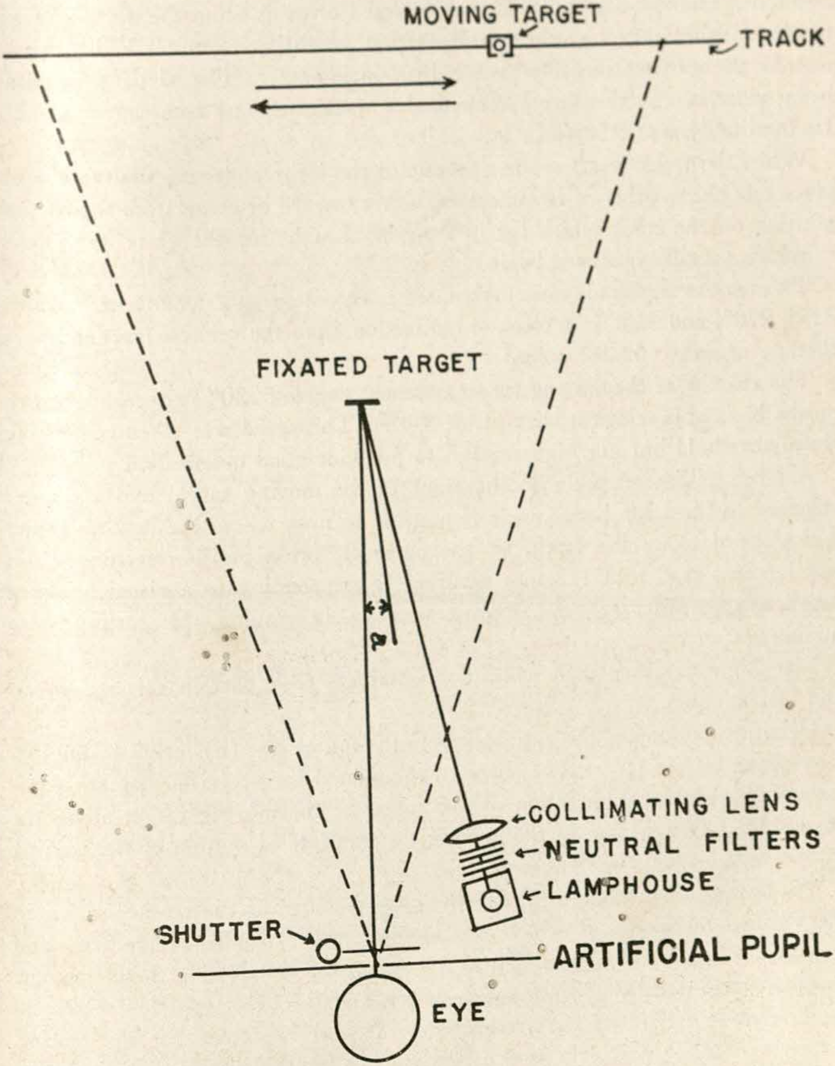

FIGURE 1
Schematic Diagram of Apparatus

time of 30 seconds or 9.9 seconds. In Experiment Two, it was 42 per cent of the fixed total trial time or 12.6 seconds.

The moving target traversed an 8-foot track which could be fixed into position for target motion in the horizontal, vertical, or the four intermediate positions. The velocity and direction (toward or away from the glare source) of the moving target was controlled by a synchronous motor. The metric size of the moving target was 1/8 inch in diameter. The viewing distance being what it was, this target subtended a visual angle of zero degree 7′ 12″. Its luminosity was 8.50 ft L.

When the track was fixed in any one of the eight positions, the target was given one or the other of two motions, either toward or away from the central position on the track which lay in the line of sight for the glare source and thus was visually obscured by it.

The various meridians were designated as zero degree, 45°, 90°, 135°, 180°, 225°, 270°, and 315° in a clockwise direction from the vertical position above the line of sight (fixated point).

The velocity of the moving target was zero degree 5′ 30″ per second, chosen on the basis of previous studies (1, 5, 8, 16). This speed was taken to be fully above threshold but not high enough to preclude good observation.

A fixed millimeter scale on the track of the moving target and a pointer attached to the lamp house made it possible to note where the moving target was stopped along the track, by the observer, for a position reading. This position was read to a 1/2 mm accuracy, corresponding to a visual angle of 35.1″ at the observer's eye. Parallax errors in reading this setting were minimized.

Five highly trained male volunteer observers took part in this study; all had 20:20 vision.

In both experiments, the variables were the following: (*a*) level of illumination of the fixated target, (*b*) shape of the fixated target (circle, square, equilateral triangle), (*c*) direction of movement of the moving target along the track, and (*d*) meridian of the moving target. Rate of motion of the moving target, as indicated, was constant.

The test procedure consisted in obtaining readings of the track position in the various meridians at which the moving target totally disappeared and at which it began to reappear. These data provided for plotting three tracings, one for zero intensity of the fixated target, and the other two for the intermediate and high intensities, respectively. It was expected that the tracing for zero intensity would lie within the tracing for the intermediate intensity and that the tracing for the intermediate intensity would lie within the tracing for the high intensity. This was, of course, expected for all three shapes of fixated targets or, in this case, both the glare and interference source.

D. RESULTS

As was expected, the fixated target was, in effect, larger than its metric dimensions, as can be seen from the tracings (dotted lines) in Figure 2. Tables 1, 2, and 3 show the differences in expected size for the three different targets, for each of the two intensities above zero. The size of the target illustrated by solid lines in Figure 2 was, in effect, its metric size. The statistical significance between the sizes is also given in the tables for one experimental data series per glare source shape.

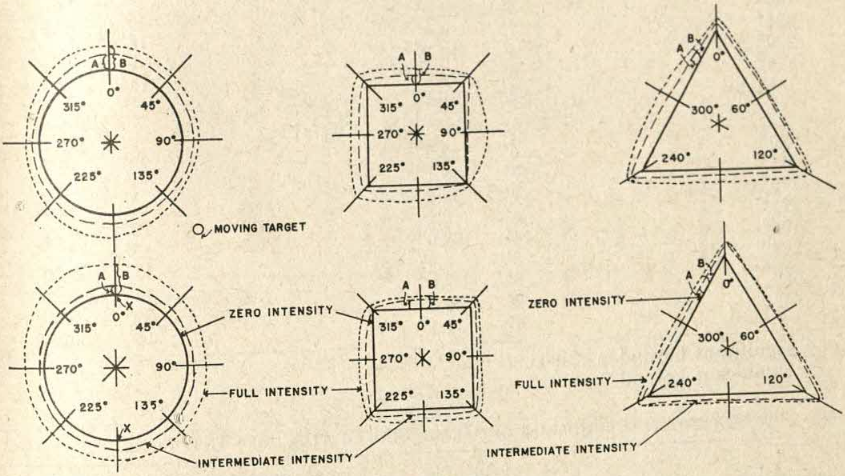

FIGURE 2
EFFECTIVE BOUNDARIES OF GLARE SOURCES HAVING DIFFERENT SHAPES
(Lower figures are the result of an entire experimental replication)

For all the meridians studied, a roughly concentric arrangement of the tracings was found. The relative size of the moving target is shown in Figure 2 as a small circle. The two levels of intensity for the fixated target differed by a ratio of 2:1 under most boundary conditions. It will be noted that there was some tendency to "fill out" or "round out" the shape of the fixated target toward a circle.

Table 4 shows the effects obtained for disappearance and reappearance position of the moving target for different boundary forms—curved, angular (60° and 90°), and straight. In all cases, the path of the moving target was either normal to the boundary (curved and straight) or bisected the angle (60° and

TABLE 1

Effective Boundaries of Glare Source at Three Intensities

Meridian	Visual angle	
	(A) .4 log Intensity	(B) High Intensity
	Experiment One (*Circle*)	
0°	10′ 3.3″*	15′ 25.4″*
45°	5′ 39.4″**	11′ 27.7″*
90°	7′ .6″*	10′ 16.7″*
135°	4′ 6.1″*	8′ 44.1″*
180°	5′ 33.9″*	13′ 48.3″*
225°	7′ 4.4″*	12′ 12.4″*
270°	5′ 13.2″*	9′ 33.4″*
315°	4′ 41.2″**	11′ 2.1″*
Mean	5′ 30.8″	11′ 35.6″*
	Experiment Two (*Circle*)	
0°	6′ 32.3″	20′ 35.9″
45°	7′ 46.1″	13′ 16.5″
90°	9′ 38.3″	12′ 39.2″
135°	6′ 9.1″	12′ 47.0″
180°	14′ 55.0″	13′ 8.0″
225°	5′ 4.4″	12′ 26.6″
270°	3′ 8.0″	13′ 8.0″
315°	2′ 26.2″	20′ 33.0″
Mean	6′ 31.1″	14′ 29.5″

* Significant beyond $p < .005$.
** Significant beyond $p < .05$.

TABLE 2

Effective Boundaries of Glare Source at Three Intensities

Meridian	Visual angle	
	(A) .4 log Intensity	(B) High Intensity
	Experiment One (*Square*)	
0°	5′ 16.3″	8′ 59.2″
45°	7′ 1.8″	7′ 57.4″
90°	0′ 37.9″	12′ 3.3″
135°	5′ 51.5″	8′ 50.0″
180°	4′ 37.6″	11′ 7.8″
225°	8′ 31.4″	14′ 9.2″
270°	3′ 59.7″	9′ 22.4″
315°	9′ 49.1″	14′ 18.3″
Mean	5′ 36.6″	10′ 31.2″
	Experiment Two (*Square*)	
0°	3′ 2.1″	5′ 55.0″**
45°	6′ 26.0″**	8′ 0.1″*
90°	2′ 0.1″	4′ 38.4″**
135°	4′ 38.4″**	7′ 54.4″*
180°	2′ 17.7″	6′ 29.4″**
225°	6′ 17.5″*	9′ 28.0″*
270°	4′ 24.2″**	8′ 14.9″*
315°	2′ 20.6″	4′ 44.0″**
Mean	3′ 29.5″	6′ 35.8″

* Significant beyond $p < .005$.
** Significant beyond $p < .05$.

TABLE 3
EFFECTIVE BOUNDARIES OF GLARE SOURCE AT THREE INTENSITIES

	Visual angle	
Meridian	(A) .4 log Intensity	(B) High Intensity
	Experiment One (Triangle)	
0°	6′ 41.4″	6′ 58.3″
60°	0′ 2.6″	8′ .8″
120°	6′ 55.4″	10′ 8.7″
180°	3′ 7.7″	7′ 15.7″
240°	9′ 37.2″	11′ 5.1″
300°	6′ 44.2″	13′ 34.8″
Mean	5′ 31.0″	9′ 22.2″
	Experiment Two (Triangle)	
0°	5′ 31.1″**	7′ 48.2″**
60°	7′ 39.7″**	7′ 37.4″**
120°	9′ 7.7″*	10′ 17.8″*
180°	3′ 37.2″	4′ 53.1″**
240°	6′ 24.0″**	6′ 56.2″**
300°	4′ 47.6″**	8′ 50.0″*
Mean	4′ 33.7″	7′ 43.7″

* Significant beyond $p < .01$.
** Significant beyond $p < .05$.

90°). It will be noted that the angular boundaries tend to produce similar effects for both levels of intensity of the fixated target, whereas the 2:1 ratio held for the high *vs.* the intermediate intensity levels for the straight and curved boundaries.

TABLE 4
EFFECTS PRODUCED BY DIFFERENT BOUNDARY FORMS UPON BOUNDARY-EDGE DETERMINATION

	Visual angle (minutes & seconds)	
Boundary form	.4 log Intensity	High Intensity
Curved (16)	6′ 26.8″	12′ 29.3″
90° (8)	5′ 36.1″	8′ 26.7″
60° (6)	6′ 41.1″	8′ 35.7″
Straight (14)	3′ 31.8″	7′ 33.8″

Note: Numbers in parentheses denote the number of sets of six IN *and* six OUT trials upon which each angle is based.

Since the traditional psychophysical method of limits was used in both experiments, it was supposed that the response times for the two sets of readings—in and out "disappearance," "reappearance"—would differ. The one task —i.e., of indicating disappearance—might be a surer one and require less time in relation to actual position of the moving target than the opposite task. One might also expect a greater statistical variance for the "appearing" trials than for the "disappearing" trials.

To test this, the variance was calculated for one experimental data series for each of the three fixated target forms. For the circular forms, only two of the 24 possible combination-of-variables cells were significant and both were in the direction opposite to expectation. For the square forms, 16 of the 24 cells exhibited larger "appearance" (OUT) variances, but only two of these were significant ($p < .01$); these two were in the expected direction. For the triangular forms, four cells out of 18 exhibited statistical significance, one in the expected direction, and three in the opposite direction.

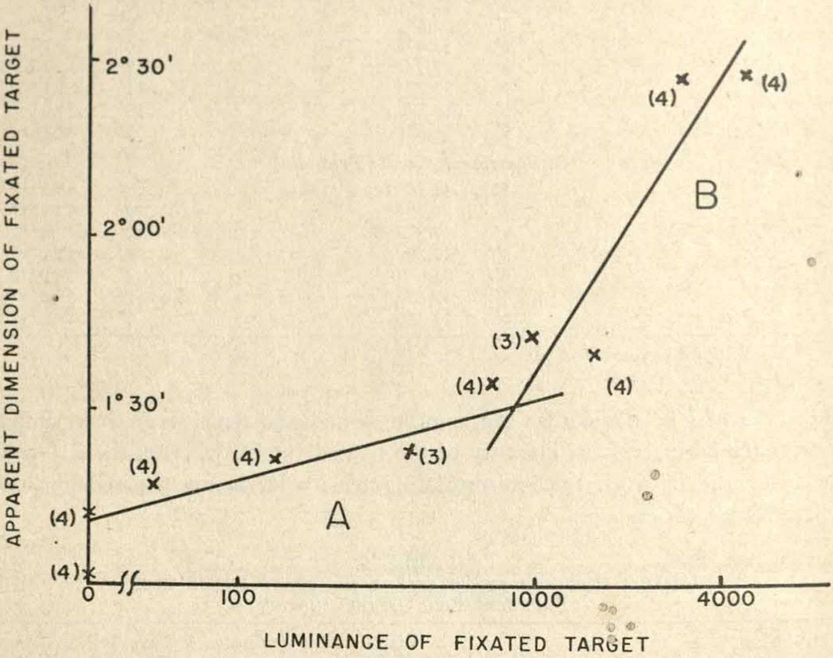

FIGURE 3
APPARENT EXPANSION OF THE FIXATED TARGET AS A FUNCTION OF ITS LUMINANCE (FT L.)

These findings do not support the idea that judgments in the OUT direction must take longer than in the IN direction.

Variance, in general, tended to be greater for each of the three target forms with the high intensity than with the intermediate or zero intensities. Variance for the various meridians did not manifest any decided trends. Despite the findings that the responses with the targets at full intensity were more variable, the subjects reported that the responses with the target at zero were the most difficult.

Experiment One was repeated in its entirety and the results are shown in the bottom diagrams in Figure 1. The mean difference between target boundary and apparent edges (tracings) for full intensity and zero intensity, in the two experiments, was under four minutes of visual angle, and under two minutes of visual angle for the intermediate intensity and zero.

Calculations of the nature of the gradient, representing the expansion in effective size (maximum dimension) of the fixated target in relation to its intensity, showed that it was not linear. Two straight lines differing in slope seemed to describe the data fairly well. These intersected at an intensity of about 1000 ft L. or a little less for the targets of the size we used (see Figure 3). Numbers in parentheses accompanying each data point represent the number of sets of 24 trials each is based upon. It is difficult to determine from present data whether this curve is asymptotic or continues its upward trend above 4000 ft L. Research with target luminance much greater than 4000 ft L. is presently underway.

E. SUMMARY

A small target moving in the frontal plane was made to disappear and reappear by passing it behind a larger fixed target of much greater luminance. The angular positions at which this disappearance and reappearance occurred were measured. These positions indicated that the fixed target was, in effect, larger than physical measurements of it would signify. Since a luminous target casts not only its expected image upon the retina but also a tapering amount of stray illumination, this stray illumination makes the target image a tapered rather than an abrupt affair and covers a greater retinal area than expected. The moving images of other less intense targets, as they approach the peripheral region of the taper of the fixed target, may be obscured just as when they fall nearly at its center. Hence, objects producing these images may be made to disappear and reappear at positions over a greater angular extent in the frontal plane than otherwise expected.

This was found to be the case in the present investigation. Three levels of intensity for the fixed retinal image were used and the positions of disappearance and reappearance were different for each, in the way expected.

REFERENCES

1. AUBERT, H. Die Bewegungsempfindung. *Arch. Ges. Physiol.*, 1886, **39**, 347-370.
2. BARTLEY, S. H. The comparative distribution of light in the stimulus on the retina. *J. Comp. Psychol.*, 1935, **19**, 149-154.
3. BARTLEY, S. H., & FRY, G. A. An indirect method for measuring stray light within the human eye. *J. Opt. Soc. Amer.*, 1934, **24**, 342-347.

4. Bartley, S. H., & Seibel, J. L. A further study of entoptic stray light. *J. of Psychol.,* 1954, **38**, 313-319.
5. Bourdon, B. *La Perception Visuelle de L'Espace.* Paris: Librairie C. Reinwald, 1902.
6. Boynton, R. M. Stray light in the human electroretinogram. *J. Opt. Soc. Amer.,* 1953, **43**, 442-449.
7. Boynton, R. M., Enoch, J. M., & Bush, W. R. Physical measures of stray light in excised eyes. *J. Opt. Soc. Amer.,* 1954, **44**, 879-886.
8. Brown, J. F. The threshold for visual movement. *Psychol. Forsch.,* 1931, **14**, 249-268.
9. De Mott, D. W., & Boynton, R. M. Retinal distribution and entoptic stray light. *J. Opt. Soc. Amer.,* 1958, **48**, 13-21.
10. Fry, G. A. A re-evaluation of the scattering theory of glare. *Illuminating Engineering,* 1954, **49**, 98-102.
11. Fry, G. A., & Alpern, M. The effect of foveal vision produced by a spot of light on the sclera near the margin of the retina. *J. Opt. Soc. Amer.,* 1953, **43**, 187-188.
12. ———. The effect of a peripheral glare source upon the apparent brightness of an object. *J. Opt. Soc. Amer.,* 1953, **43**, 189-195.
13. ———. Effect of a peripheral glare source upon the apparent brightness of an object. *Illuminating Engineering,* 1955, **50**, 31-38.
14. Fry, G. A., & Bartley, S. H. The relation of stray light in the eye to the retinal action potential. *Amer. J. Physiol.,* 1935, **111**, 335-340.
15. Geldard, F. A. Brightness contrast and Heyman's law. *J. Gen. Psychol.,* 1931, **5**, 191-206.
16. Grim, K. Uber die Genaugkeit der Wahrnemung und Ausfuhrung von Augenbewegungen. *Z. f. Sinnesphysiol.,* 1910, **45**, 9-26.
17. Haines, R. F. The effect of high luminance sources upon the visibility of point sources. *Advances in the Astronautical Sci.,* 1965, **20**, in press.

National Aeronautics and Space Administration
Ames Research Center
Moffett Field, California 94035

Published as a separate and in *The Journal of Psychology,* 1966, **62**, 267-278.

LOUDNESS ADAPTATION AS A FUNCTION OF FREQUENCY, INTENSITY, AND TIME*[1]

Department of Psychology, Tufts University; and Autonetics

WARREN H. TEICHNER AND ERNEST SADLER

A. INTRODUCTION

The decline in loudness of pure tones under sustained stimulation has received considerable attention in the past few years. Although they have not supported their conclusions with specific experimental citations, some reviewers (9, 17) have noted that loudness probably does not adapt with continued stimulation.

In 1929, von Bekesy measured the decrease in loudness of tones of various frequencies and intensities and found that it varied greatly as a function of the intensity of the stimulating tone but bore only a slight relationship to frequency. Presenting a long-continued tone to one ear and a second tone of 0.2-second duration to the other ear just after the first tone ended, von Bekesy instructed the subjects to adjust an attenuator so as to produce the same loudness in the second tone as the terminal portion of the first one. He reported that as the duration of the first tone was increased, the intensity of the second tone had to be decreased. Because of the extremely short duration of the second tone, it seems unlikely that the subjects actually manipulated an attenuator to equate the two tones for loudness. Rather it would seem that this manipulation was done by the experimenter and that the subjects gave verbal or other signalling responses. To permit recovery between test runs, von Bekesy found it necessary to follow each presentation with a quiet period of 3 to 8 minutes depending on the duration of the first one.

In 1930, using essentially the same technique, Wood (16), as described in Geldard, also found that as the duration of the first or standard tone increased, the intensity level of the comparison tone necessary to provide a loudness match decreased. Hood (4), in 1950, described a technique per-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 7, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was carried out at the University of Massachusetts, supported by Grant AFOSR 60-2 with the Air Force Office of Scientific Research on which the senior author was principal investigator. The paper in its present form was prepared by the senior author at the Guggenheim Center for Aerospace Health and Safety, Harvard University, School of Public Health under contract No. AF-AFOSR-958-65 between the Air Force Office of Scientific Research and Harvard University.

mitting successive measurements which differed from that of both von Bekesy and Wood in that the two tones were on simultaneously. In Hood's method, simultaneous tones were first presented to each ear for a brief period (10 to 20 seconds). The intensity of one tone, the standard, was fixed, while the intensity of the other tone, the comparison, was variable. The subject attenuated the comparison stimulus until the two tones appeared equally loud. Two or three matches were made in this manner. Then without interruption the standard tone alone was continued while the comparison tone was presented only at periodic intervals for short durations during which the subject again equated the tones for loudness. The difference between the intensity level required for a loudness match initially and the intensity level required after the standard tone had been on for some time was taken as the amount by which the response of the test ear had declined under sustained stimulation. This decline has been called "perstimulatory fatigue" or "adaptation" by various investigators.

Using this procedure, Hood measured adaptation for various intensities and frequencies. He concluded that adaptation was essentially complete in about 3.5 minutes, irrespective of intensity, and that the maximum amount of adaptation increased as a function of the intensity of the standard tone. At high intensities, adaptation amounted to as much as 50 db.

In view of the considerable degree of adaptation shown by Hood's results, Egan (2) attempted to determine if the amounts of adaptation were a function of the method used to equate the tones for loudness. He compared an equal-loudness method, in which the subject was instructed to make the tones equally loud, and the median-plane localization method, in which the subject was told to adjust the sound until only one tone was heard, this being localized in the median plane. He had subjects match identical frequencies and also match frequencies that differed from one another by 5 to 200 cps. The frequency difference was expected to prevent the subject from localizing the sound in the median plane, and thus force him to make a true equal-loudness balance. Egan concluded from the results that the two methods did not differ in the amount of adaptation measured. He also found greater amounts of adaptation for higher frequencies.

Egan and Thwing (3) found that considerably less adaptation was measured by von Bekesy's method, in which the comparison tone comes on after the standard tone has ceased, than by Hood's method, in which the loudness balance is made while both tones are on.

Jerger (6), using the median-plane localization method, extended the study of adaptation to pure tones by using a wider range of frequencies and intensities than had been used before. In view of Egan's (2) results, it is of some interest that at high frequencies (2000 to 8000 cps) a number of Jerger's subjects were unable to localize the stimulus in the median plane during the period of sustained stimulation. In these cases, they were told to make an equal-loudness match. Jerger found that the time at which maximum adaptation occurred was a function of both frequency and intensity. At frequencies up to 500 cps, maximum adaptation seemed to be reached after 3 minutes of stimulation, but in the range from 1000 to 8000 cps, only adaptation with the very low intensities showed any tendency to stabilize at a maximum value. At the higher intensities, 70 to 90 db, adaptation continued to increase even after 5 minutes. This result is not in agreement with Hood's conclusion that the point of maximum adaptation is reached at about 3.5 minutes, irrespective of intensity.

In another study by Jerger (1956), adaptation was measured for a 5000 cps tone at 80 db over a 10-minute period of sustained stimulation. The point of maximum adaptation was not reached until the 7th minute, a finding in close agreement with Carterette's (1) results for adaptation to sustained thermal noise, and removed from Hood's (4) finding of 3.5 minutes.

There is also wide variation as to the maximum amount of adaptation which occurs. Most investigators (2, 3, 6, 13, 14) reported between 20 to 25 db while Hood (4) reported up to 50 db and Palva (10), using a von Bekesy type attenuator, only 8 db.

Small (11) points out that a likely cause of these wide differences lies in the wide variation in psychophysical procedures which have been used as well as in other methodological techniques. The purpose of this study, therefore, was to reinvestigate the question by designing an experiment intended to overcome possible artifacts resulting from procedures of previous studies.

Perhaps the most basic point among studies reporting adaptation is the question of whether the subject actually matched the tones for loudness or whether his response was due to some other kind of judgment. Egan (2) found no difference in amount of adaptation reported when the subject was instructed to match loudness and when he was instructed to localize in the median plane. However, Jerger's (6) finding that some of his subjects were unable to localize in the median plane suggests the possibility of ambiguous reporting in at least some of Egan's subjects. The finding that subjects were able to match tones of different frequency suggests that they can make a loudness match, but not that they actually do so. Further indication that the subject is in fact matching loudness might be gained by requiring the subject to make judgments of loudness ratios as well as equal loudness judgments.

If he is able to do both and if both show adaptation, then considerable support might accrue to the genuineness of loudness as his judgmental response and to the adaptation phenomenon as well. This was the technique of the present study, to use equal-loudness and twice-as-loud judgments.

A second point concerns the similarity of the standard and comparison tones during the period when the subject makes his judgment. It would seem reasonable that the greater the physical difference between these two, the more difficult the subject's judging task would be. Thus, the most acceptable procedures would seem to be those which minimized the differences between tones. The two tones should not only be identical in frequency, but also as nearly alike as possible in both physical and apparent intensity. Thus, the common procedure of presenting a highly attenuated comparison tone may permit the possibility of confounding physical and subjective intensity in the subject's judgment at the instant of stimulation as well as successively during the period of adjustment. If adaptation is a function of intensity then this procedure would have the advantage of minimizing adaptation of the test ear during the adjustment period. On the other hand, it has a disadvantage in that it requires the subject to compare successive pairs of tones of widely different subjective intensity. Moreover, one might expect that the time required for the adjustment would also be longer than if the tone were presented at its original intensity and to the extent that adaptation depends on time the test ear itself might adapt considerably. In fact, adjustment times of 15 seconds have been the most widely used. Related disadvantages of this procedure are that the subject is likely to overshoot in making his adjustment and if a step-attenuator is used, as has usually been the case, possibly important amounts of overshooting are difficult to avoid. Further, rather than acting as a control, the procedure of randomly attenuating the comparison tone before each match (2, 3, 6, 13) might complicate the situation by introducing a variable reaction time in the subject's judgment (12).

Little use has been made of the procedure of presenting the subject with a comparison tone identical in intensity (and frequency) to the standard. The primary objection to this procedure is that if adaptation is a function of intensity, such a procedure might be expected to lead to more rapid adaptation of the test ear; and, consequently, unless very rapid judgments could be made, this should result in less apparent adaptation than would be obtained by presenting an attenuated comparison tone. However, such a procedure might permit extremely rapid judgments, and if the subject used a continuously variable attenuator he could make rapid adjustments of the comparison tone. The latter consideration led to the adoption of this technique in the present experiment.

Another important experimental point concerns the difference in sensitivity of the two ears. As far as can be determined, no previous investigator has reported using subjects with ears matched in sensitivity. The usual procedure has been to use the difference in intensity obtained from an initial loudness match and one obtained after a period of continued stimulation. This procedure is acceptable only on the assumption that ears of different sensitivity adapt along the same time function. In view of the disagreement among studies reported, a safer procedure would appear to be to select subjects, in advance, with equally sensitive ears. This procedure, which was taken in the present study, permitted direct use of the difference between the standard and comparison ears at any point in time.

Also of importance is the duration of the recovery interval between stimulations. This has varied from one experiment to another, from 20 seconds to 8 minutes depending on the duration of the standard tone. Because of the uncertain nature of loudness adaptation, the length of the interval required for recovery would appear to be critical. Some consideration was given to this problem by running a preliminary experiment in which subjects were placed under the combinations of frequency and intensity used in the present study. Stimulation periods were 0.25, 0.50, 1.0, 2.0, and 5.0 minutes and the corresponding recovery times were 1.0, 1.0, 3.0, 5.0, and 10.0 minutes. The results obtained under these conditions did not differ from the results obtained when the recovery times were equal to the stimulation period plus 30 seconds. Therefore, times equal to the stimulation period plus 30 seconds were used in the present study.

B. METHOD

1. *Subjects*

The subjects were four female undergraduate students. These students were selected from 35 volunteers. The basis for selection was an agreement between the two ears within 2 db at absolute threshold of the frequencies used in the experiment. The subjects were paid for their participation. These subjects had previously participated in an experiment of 3-month duration which involved continuous binaural stimulation and had also been given pre-experimental practice in the judgments required for the present study. Thus, they were considered to be highly skilled subjects at the start of the present study.

2. *Apparatus*

The signal was generated by a Hewlett-Packard audio oscillator and led to a step-up audio transformer. At this point the signal was divided into a standard and comparison channel. The channel for the standard tone went to the experimenter's continuously variable attenuator (General Radio), and from there directly to one earphone. The comparison channel was passed through an Atcotrol timer to a 2 db step attenuator (General Radio), then to the subject's continuously variable attenuator (General Radio). At this point it was further divided with one channel going to the subject's other earphone and another channel to a General Radio Graphic Level Recorder. PDR-10 earphones, mounted in MX-41/AR cushions, were matched in impedance; and all impedances within the system were appropriately matched. Calibrations were made with a Hewlett-Packard vacuum-tube voltmeter, the Graphic Level Recorder, and a General Radio Sound Level Meter. The subjects sat in a sound-absorbent chamber; the experimenter monitored the experiment from an outside control area permitting observation through a one-way window.

3. *Procedures*

Prior to each experimental session, the subject was required to rest in the sound room for 5 minutes before donning the earphones. The signal for donning the earphones and for other phases of the session was a silent flashing light controlled by the experimenter.

Two frequencies, 3500 cps and 5000 cps, were used, each at two intensities, 85 and 95 db, re 0.0002 dynes per square centimeter. Each of these four combinations was presented to the standard ear for 0.25, 0.50, 1.0, 2.0, and 5.0 minutes, and then was also presented to the test ear for judgment. During judgment periods, therefore, both ears were stimulated simultaneously. For each of the 20 intensity-frequency-time combinations, the subject made an equal-loudness and a twice-as-loud judgment, setting the attenuator as rapidly as possible. These 40 observation conditions made up an experimental series. The experimental series was replicated three times for each subject, with each replication consisting of a different random presentation of the combinations. Sessions were limited to one hour, with no more than two sessions for one subject in any one day. A complete replication took approximately five sessions.

The following is an example of an actual run: a 5000 cps tone at 95 db was transmitted to one ear (standard tone). After 2 minutes the same tone was also transmitted to the other ear (comparison tone). Immediately upon hearing the comparison tone, the subject, by means of his attenuator, manipulated the intensity level of the comparison tone until it sounded equal in loudness to the standard tone. Upon completion of the match, both tones were discontinued by the experimenter. A continuous graphic record was kept of the subject's attenuator settings. The interval between runs was equal to the period of stimulation by the standard tone plus 30 seconds.

C. RESULTS

The amount of adaptation in db, re 0.0002 dynes per square centimeter, for the equal-loudness matches, was obtained as the difference between the

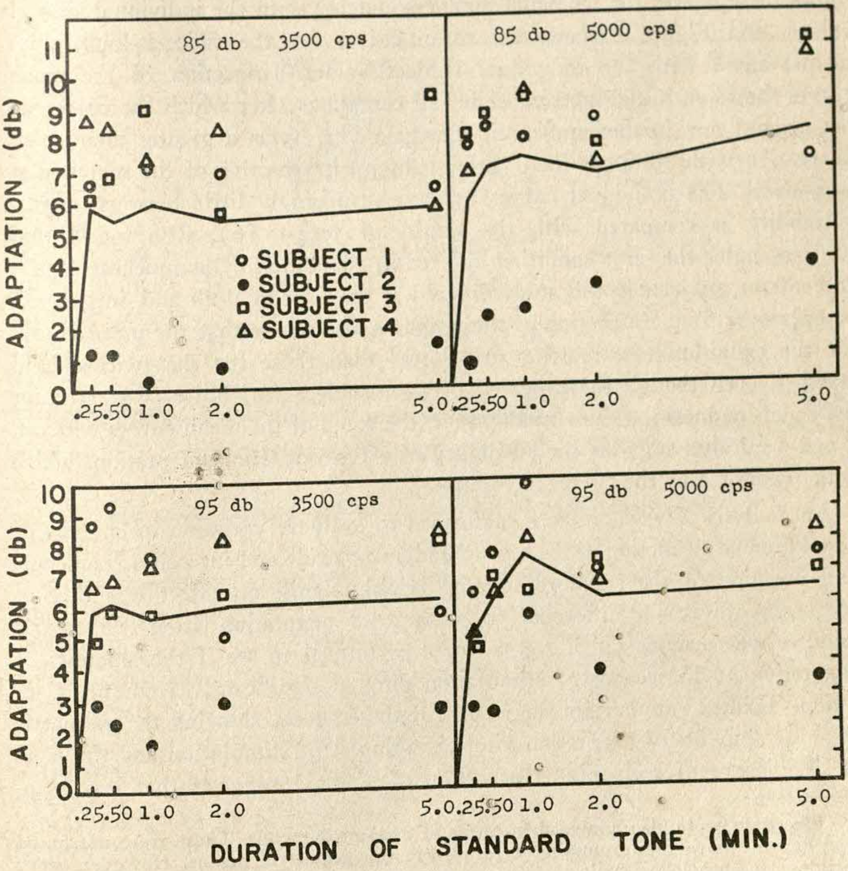

FIGURE 1
MEAN AND INDIVIDUAL ADAPTATION FOR EQUAL-LOUDNESS JUDGMENTS

final intensity setting of the comparison tone and the intensity of the standard tone. A less direct method was necessary in the case of the twice-as-loud judgments. The sone values of the standard tones were computed and then doubled. These doubled sone values were then reconverted to db. The final intensity settings of the comparison tones were then subtracted from the reconverted db values to give the amount of adaptation. The intensity values of the comparison tones used in computing adaptation were the geometric means of the intensity settings of the three replications.[2]

A mean adaptation curve was derived for each experimental condition by obtaining the geometric mean of all the subjects at each condition. Figure 1 shows these curves for the equal-loudness matches with the individual subject values, and Figure 2 shows the mean curves for the twice-as-loud judgments, again with the individual subject values. Inspection of the mean curves shows an initial increase under all conditions after which the curves do not suggest any further systematic changes. There was a greater amount of adaptation with 5000 cps than with 3500 cps, irrespective of the match that was made. The individual curves indicate surprisingly little between-subject variability as compared with the results of Jerger (6), who found that subjects under the same condition differed in adaptation by as much as 40 db. In contrast, subjects in this study differed at the most by 9 db and, in general, by only 4 or 5 db. Inspection of these figures also shows that the mean curves for the equal-loudness matches are flatter than those for the twice-as-loud matches, even though the range of between-subject variability is greater for the equal-loudness matches. In Figure 2, the trend of the mean curves between 2 and 5 minutes suggests the possibility of a recovery from adaptation, which is not exhibited by the curves in Figure 1.

An analysis of variance was performed to evaluate variation in amount of adaptation as a function of the type of judgment made and intensity, frequency, and duration of the standard tone. The only significant effect was that of frequency. Since no difference in amount of adaptation from .25 to 5.0 minutes was demonstrated, *t* tests were performed to see if the amounts of adaptation at 15 seconds of stimulation differed significantly from no adaptation. Table 1 summarizes the results of these *t* tests, showing the geometric mean of amounts of adaptation after .25 minutes of stimulation and the level of significance of each mean. Inspection of Table 1 suggests that less adap-

---

[2] The data were also analyzed in terms of arithmetic means. These were not found to differ by more than rounding errors from the geometric means. However, since decibles are on a logarithmic scale, geometric means are reported. It should be noted, though, that the more variable the subject is in his repetitions of the same judgment, the more greatly the arithmetic and geometric means may differ. Thus, their agreement in the present study reflects the favorable reliability of the data.

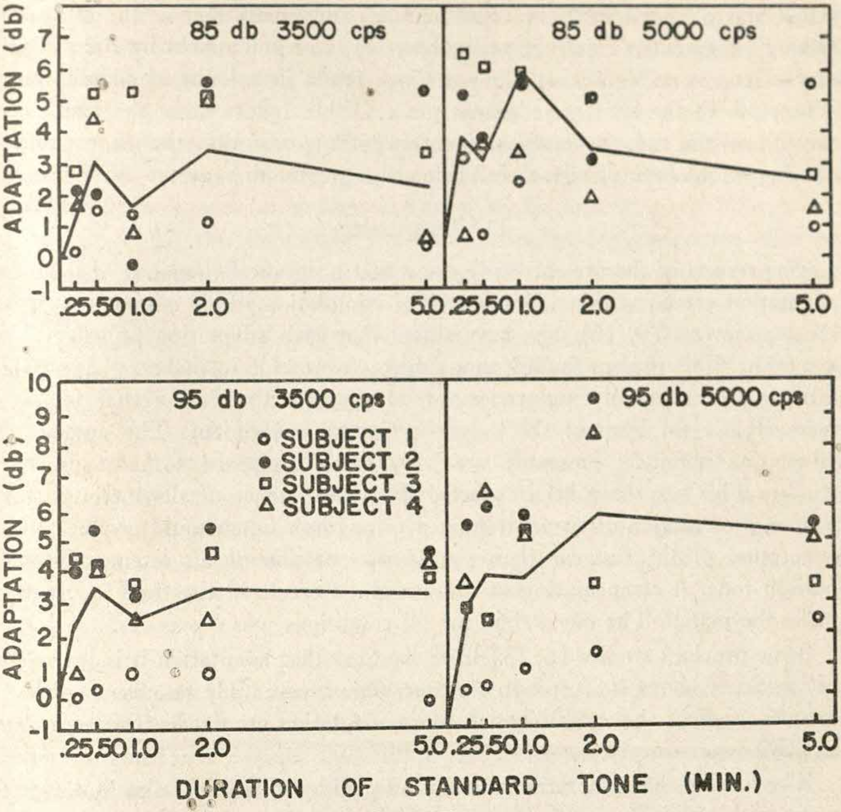

FIGURE 2

MEAN AND INDIVIDUAL ADAPTATION FOR TWICE-AS-LOUD JUDGMENTS

TABLE 1

TABLE OF *t* TESTS BETWEEN DIFFERENCES IN ADAPTATION BEFORE STIMULATION AND AFTER .25 MINUTES OF STIMULATION

Judgment made	Stimulus condition	Geom. mean adaptation (db)	t
equally loud	85 db 3500 cps	5.66	3.56**
equally loud	85 db 5000 cps	5.94	3.41**
equally loud	95 db 3500 cps	5.82	4.69***
equally loud	95 db 5000 cps	4.70	6.02***
twice as loud	85 db 3500 cps	1.64	2.93*
twice as loud	85 db 5000 cps	3.79	3.03*
twice as loud	95 db 3500 cps	2.36	2.07
twice as loud	95 db 5000 cps	3.08	3.35**

* $p < .05$.
** $p < .025$.
*** $p < .01$.

tation was obtained with the equal-loudness judgments than with the twice-as-loud judgments. This suggestion, however, was not upheld by the analysis of variance as no significant difference was found in amount of adaptation as a function of the kind of judgment made. Table 1 does show that with one exception—the twice-as-loud judgment at 3500 cps, 95 db—the mean amount of adaptation obtained at each condition was greater than zero.

D. Discussion

The results of the present study show that a significant amount of loudness adaptation occurs within .25 minutes of stimulation. This does not support those reviewers (9, 17) who have stated that such adaptation probably does not occur. The present finding that subjects were able to make reliable judgments of both equally and twice-as-loud supports the notion that loudness was actually the basis of the judgment in this experiment. The amount of adaptation obtained, however, was very small compared to most previous studies. This may have been expected as a consequence of the method used; that is, presentation of an unattenuated comparison tone should produce more adaptation of the test ear than should presentation of an attenuated comparison tone. A compensating factor was the decrease in the time required to make the match. The mean time for all conditions was 5 seconds.

Some previous studies (6, 13) have reported that adaptation is independent of frequency above 1000 cps. In contrast, the present study found a significant frequency effect above 1000 cps as more adaptation occurred at 500 cps than at 3500 cps.

The finding that adaptation was not affected by intensity is also in disagreement with previous results. This too might have been due to the use of the unattenuated test tone. That is, if adaptation were complete within 15 seconds, differences in adaptation produced by 10 db differences in stimuli might be difficult to measure. Either a stimulus duration of less than .25 minutes or a difference in stimulus intensity of more than 10 db might result in a significant intensity effect. This problem deserves further study.

As an explanation for the results of this experiment, the possibility exists that contraction of the tympanic muscles may be partially or wholly responsible for the adaptation obtained. Luescher (7), in his observations on the stapedius muscle, reported an increase in duration of muscle contraction with an increase in intensity of the stimulating tone, with the high intensities giving a steady contraction that lasted throughout the presentation of the tone up to 10 or 15 seconds. This agrees well with the maximum adaptation time obtained in the present experiment.

Metz (8) found that these contractions rapidly attained a maximum, remained at maximum for several seconds, and then, although still being stimulated, slowly relaxed. This process appears to be a reflex adaptation and not a muscular fatigue, for Metz found that if the tone were stopped and then immediately presented again, the contraction reappeared with full magnitude. Jepsen (5) reported that these muscle reflexes occurred only when the stimulating tone was at a high intensity level. According to Wever and Lawrence (15), the amount of protection afforded by contraction of these muscles varies greatly with frequency, with the greater protection being afforded for the lower frequencies. This finding is in opposition to that of the present study in that more adaptation was obtained at 5000 cps than at 3500 cps. On the other hand, the findings of Luescher (7), Metz (8), and Jepsen (5) do not provide specific disagreement with the lack of significant intensity effect found in the present study. That is, the 10 db difference between the tones used may be within the difference limen of the tympanic muscles. However, as noted above, that is a question deserving further study. The relationship between the response of the tympanic muscles to continued stimulation and the present data is certainly suggestive. However, as noted by Small (11), a variety of theoretical approaches are possible.

E. Conclusions

The method used did not provide results in agreement with previous studies. In contradiction to these studies, intensity did not affect the amount of adaptation obtained; frequencies above 1000 cps did affect adaptation; adaptation was complete within .25 minutes and the amount of adaptation obtained was relatively small. It was demonstrated, however, that loudness was the basis of the judgments made and that the amount of adaptation is independent of whether the subject makes equal-loudness settings or double-loudness settings.

References

1. Carterette, E. C. Perstimulatory fatigue for continuous and interrupted noise. *J. Acoust. Soc. Amer.*, 1955, **27**, 103-111.
2. Egan, J. P. Perstimulatory fatigue as measured by heterophonic loudness balances. *J. Accoust. Soc. Amer.*, 1955, **27**, 111-120.
3. Egan, J. P., & Thwing, E. J. Further studies on perstimulatory fatigue. *J. Acoust. Soc. Amer.*, 1955, **27**, 1225-1226.
4. Hood, J. D. Studies in auditory fatigue and adaptation. *Acta Otolaryng.*, 1950, Suppl. No. 92.
5. Jepsen, O. The threshold of the reflexes of the intra-tympanic muscles in a normal material examined by means of the impedance method. *Acta Otolaryng.*, 1951, **39**, 406-408.
6. Jerger, J. F. Auditory adaptation. *J. Acoust. Soc. Amer.*, 1957, **29**, 357-363.

7. Luescher, E. Die Funktion des Musculus Stapedius beim Menschen. *Z. Hals-Nasen-Ohrenheilk.*, 1929, **23**, 105-132.
8. Metz, O. Studies on the contraction of the tympanic muscles as indicated by changes in the impedance of the ear. *Acta Otolaryng.*, 1951, **39**, 397-405.
9. Osgood, C. E. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.
10. Palva, T. Studies on per-stimulatory adaptation in various groups of deafness. *Laryngoscope*, 1955, **65**, 829-847.
11. Small, A. M., Jr. Auditory adaptation. In J. Jerger (Ed.), *Modern Developments in Audiology*. New York: Academic Press, 1963.
12. Teichner, W. H. Recent studies of simple reaction times. *Psychol. Bull.*, 1954, **51**, 128-149.
13. Thwing, E. J. Spread of perstimulatory fatigue of a pure tone to neighboring frequencies. *J. Acoust. Soc. Amer.*, 1955, **27**, 741-748.
14. von Bekesy, G. Experiments in Hearing. New York: McGraw-Hill, 1960.
15. Wever, E. G., & Lawrence, M. Physiological Acoustics. Princeton: Princeton Univ. Press, 1954.
16. Wood, A. G. Unpublished Master's thesis, University of Virginia, Charlottesville, 1930 (cited in Geldard, F. A., *The Human Senses*). New York: Wiley, 1953.
17. Woodworth, R. S., & Schlosberg, H. Experimental Psychology. New York: Holt, 1954.

Department of Psychology
Tufts University
Medford 55, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 279-285.

CATEGORY CLUSTERING IN FREE RECALL*[1]

Department of Psychology, University of Louisville

JOHN A. ROBINSON

A. INTRODUCTION

Typically, studies of clustering have involved only one presentation and one recall of the list. While variations in the composition of the list (1) and its method of presentation have been explored (2), little attention has been given to the potential value of treating such materials as a problem of multitrial list learning. There are many issues which might be clarified by such experiments, and in this report we shall discuss two of these: category clustering and the role of the category name in recall.

When a categorizable word list is presented several times and successive recalls of the list obtained, we might expect that an *S*'s tendency to order his recall, as reflected in intracategory item clustering, would be extended to the categories as well. That is, *S*s should tend to group categories of items together in recall as well as the items within the categories. This result would particularly be expected if we take the view that ordering assists recall and is not simply an artifact of it [e.g., (6, 7)].

While category clustering would be of some interest in itself, its occurrence would raise additional questions about the exact mechanisms by which such ordering is produced in free recall, for the mechanisms underlying category clustering are potentially more diverse than those underlying item clustering. In particular, associations between category names (either pre-existing or established by *S* during the course of learning) should be expected to play a major role in producing what might be called second-order structure in recall. The present study was designed to obtain information about the occurrence of category clustering and the processes by which it is achieved. The data reported below bearing on these two problems are part of a larger study investigating some characteristics of clustering in a multitrial design.

---

* Received in the Editorial Office, Provincetown, Massachusetts, March 8, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by grants from the University of Louisville Research Council and the Arts and Sciences Research Committee. The author wishes to acknowledge the assistance of Jerry Cartwright in collecting the data.

B. Method

1. *Subjects*

The *S*s were 89 students from introductory psychology courses at the University of Louisville.

2. *Stimulus Materials and Procedure*

The list to be learned consisted of 30 items representing 10 different categories. The items were selected from 10 nonexhaustive categories prepared and used by Cohen (3). The list was recorded on tape at the rate of one item per second for presentation to *S*s. Five randomizations of the list were prepared, such that no two items from the same category occurred successively. *S*s were run in groups of five to 15. The instructions read by *E* indicated that this was an experiment on learning and remembering verbal material. The list was presented five times with a three-minute recall period following each presentation. All *S*s were given a five-page booklet to use for recall and *E* ascertained that each *S* turned to the next blank page of the booklet before beginning the next trial.

C. Results

Recall protocols were scored for (*a*) number of items correctly recalled on each trial, (*b*) number of categories recalled on each trial, and (*c*) amount of item clustering on each trial. These data are presented in summary form in Table 1. Regular increases on each of these measures were obtained with each additional trial.

Brief mention of the item-clustering index (ICI) is appropriate inasmuch as the ratio of repetition (RR) typically used as an index of item clustering was found to be unsatisfactory for present purposes. The RR fails to take account of the possibility of perfect clustering and the gaps in repetition represented by category transitions in such cases. The ICI, expressed as Ro/Co (n-1), corrects this deficiency. It is the ratio of the number of obtained intracategory item repetitions to the number of possible repetitions. The denominator is calculated by multiplying the number of categories represented in the recall protocol by n-1, the number of items per category minus one. In this experiment the quantity (n-1) was always two. Category recall was scored if at least one item from the category was contained in the recall protocol.

The protocols of 20 *S*s whose recall was characterized by perfect clustering on trials four and five provided the data for analyses of category clustering

TABLE 1
SUMMARY OF RESULTS

Variables	Trial 1 Measures		Trial 2 Measures		Trial 3 Measures		Trial 4 Measures		Trial 5 Measures	
	$\overline{X}$	*SD*	$\overline{X}$	*SD*	$\overline{X}$	*SD*	$\overline{X}$	*SD*	$\overline{X}$	*SD*
Number items recalled	12.33	2.83	19.60	4.29	23.74	3.24	26.02	4.03	27.33	3.35
Number categories recalled	6.77	1.49	8.61	1.32	8.95	1.37	9.47	.80	9.53	1.49
Item-clustering index	.24	.28	.46	.34	.63	.27	.76	.25	.82	.23

(CC). CC, if it occurred, should be most easily discerned in recall exhibiting a high degree of other types of ordering. If CC could not be demonstrated under these conditions, it would indicate that examination of less highly ordered recall data than that obtained from the 20 *S*s with perfect clustering in trials four and five would probably be unprofitable.

	C_o	M_a	T_i	E_l	G_a	C_r	T_r	W_e	A_n	M_{us}
C_o		1			2					
M_a	2						1			
T_i							1	3		
E_l		1	1		1					
G_a				1		1	1			
C_r			2						1	
T_r									2	1
W_e				1						2
A_n				1		1				
M_u						1	1			

FIGURE 1

An Example of a Category-Transition Matrix

The initial step in the analysis consists of coding the recall data into sequences of category transitions. The protocols for trials three, four, and five were coded in this manner for each of the 20 *S*s. Category transition matrices (CTM) were then prepared for each *S*. This matrix contains the enumeration of how often each category succeeded every other category in recall on trials three to five. An example of a CTM is presented in Figure 1.

Each matrix was then reduced in order to compute a category-clustering index (CCI). The reduction of the CTM consists of taking each cell where there is an entry and counting how often that pair of categories occurred together on trials three to five in both a "forward" and "backward" order. For example, by reference to the CTM of Figure 1 it can be seen that the category COUNTRIES (Co) was followed by the category CLOTHING MATERIALS (Ma) once, and preceded by this same category twice. On the other hand, the category WEATHER (We) was followed by the category MUSICAL INSTRUMENTS (Mus) twice, but never preceded by it.

The CCI is the ratio of the number of category pairs occurring contiguously (i.e., in both forward and backward contiguity) two or more times to the total number of different category pairs. Reduction of the CTM in Figure 1 would yield only sixteen different pairs of categories occurring contiguously one or more times in recall. However, nine of these pairs occur successively two or more times in recall. Thus, the CCI for this *S* would be .56.

TABLE 2
CLASSIFICATION OF CATEGORY-TRANSITION PROCESSES OBTAINED FROM ANALYSIS OF LINKING ITEMS

Class		Example	Relative frequency	Per cent
I. Forward interitem associations		England-Year England-Year	16	15.23
II. Forward category-item or item-category associations	A.	Flute-Snow Trumpet-Snow	14	13.33
	B.	Corduroy-Helium Corduroy-Sodium	14	13.33
III. Forward intercategory associations		Tennis-Day Golf-Week	13	12.38
IV. Symmetric intercategory associations	A.	Leg-Flute Trumpet-Leg	23	21.90
	B.	Denim-Oak Willow-Tweed	25	23.80

The CCI's obtained for these 20 *S*s ranged from .14 to .56 with a mean of .30 and an $SD = .12$. At present, no satisfactory way is available for determining what value of CCI would be expected by chance, so that we cannot attach any significance value to the mean CCI. However, an average redundancy of 30 per cent in the order of category recall would seem to be fairly substantial.

Some evidence for the occurrence of CC having been obtained, additional analyses were carried out in an attempt to delineate the processes underlying CC. Only those category pairs which occurred contiguously two or more times in recall were examined to see if the items linking the two categories—i.e., the terminal item of one category and the initial item in the category succeeding it—were invariant from trial to trial. Impressions gained during the CC analysis seemed to indicate that in a few cases the linking items would always be the same, but in many others they would not. Both impressions were confirmed. On the basis of this analysis of linking items, a four-fold classification of presumed category-transition processes was developed. This scheme is detailed in Table 2. The pairs of items given as examples are those which linked their respective categories on trials four and five.

D. Discussion

The occurrence of category clustering indicates that with amenable materials *S*s can establish highly structured "plans" for remembering (5). The two levels of structure identified in this experiment—intracategory item clustering and category clustering—reflect the hierarchical organization of associations which constitute such "plans." That many different kinds of associations may be involved in the organization of recall is indicated by the analysis of linking items summarized in Table 2.

Perhaps, the most interesting implication of that analysis is that category names do play a role in ordering recall. Class III and Class IV-B associations would seem to offer the strongest support for this assertion. Consider the illustration of Class III associations given in Table 2: the category GAMES was followed by the category TIME on trials four and five, yet the items linking these two categories were different on each trial. In the case of Class IV-B associations, not only are the linking items different on each trial, but, in addition, the category order is reversed. In both cases, it is necessary to postulate associations between some generic representations of the respective categories, presumably the category names, in order to account for the repeated contiguity of these categories in recall. It should be noted, too, that instances of Class III and Class IV associations comprise more than half of all the linkages analyzed.

That other processes may be involved in ordering recall is evident as well. Class I associations, for example, represent category ordering achieved by direct association of items in the respective categories; and those of Class II, while nominally identified as category-item or item-category associations, may be simply mediated associations. As an illustration, consider the following pair of linking items: FOG-TRUMPET, FOG-FLUTE. If we suppose that HORN occurred as an association to FOG, the transition from the category of weather to the category of musical instrument could be accomplished via HORN as a mediating link. We could further suppose that on the next trial FOG was followed by HORN; but due, perhaps, to response competition TRUMPET was replaced by FLUTE. Which interpretation of Class II associations is correct is a matter for further research.

It should be kept in mind that all of the analyses reported here were carried out on recall protocols exhibiting perfect item clustering on trials four and five. Factors which affect item clustering are probably maximally operative in the earlier trials so the classification of associative processes presented here may not be relevant for understanding item clustering. In particular, it should be emphasized that, although we have presented presumptive evidence for the role of category names in ordering *category* recall, what role they may have in ordering *item* recall is still unclear (1, 4).

E. SUMMARY

A list of 30 nouns drawn from 10 nonexhaustive categories was presented orally to 89 *S*s five times. A three-minute written recall period succeeded each presentation of the list. The data were scored for (*a*) number of categories recalled on each trial, (*b*) number of items correctly recalled on each trial, and (*c*) amount of item clustering. The protocols of 20 *S*s who showed perfect item clustering in recall on trials four and five were analyzed for evidence of category clustering. Category-transition matrices were prepared for each of these *S*s and the number of recurrent, contiguous, category pairs was determined. Examination of the specific items linking the categories yielded a four-fold classification of presumed category-transition processes. Each of these processes was briefly discussed.

REFERENCES

1. COFER, C. N. On some organizational characteristics of free recall. *Amer. Psychol.,* 1965, **20**, 261-272.
2. COFER, C. N., BRUCE, D., & REICHER, G. M. Clustering in free recall as a function of certain methodological variations. Technical Report No. 3, for NONR 656 (30), Pennsylvania State University, University Park, Pennsylvania, September, 1964.
3. COHEN, B. H. An investigation of recoding in free recall. *J. Exper. Psychol.,* 1963, **65**, 368-376.
4. DALLETT, K. M. Number of categories and category information in free recall. *J. Exper. Psychol.,* 1964, **68**, 1-12.
5. MILLER, G., GALANTER, E., & PRIBRAM, K. Plans and the Structure of Behavior. New York: Holt, 1960.
6. TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.,* 1962, **69**, 344-354.
7. ———. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychol. Rev.,* 1964, **71**, 219-237.

Department of Psychology
University of Louisville
Louisville, Kentucky 40208

Published as a separate and in *The Journal of Psychology*, 1966, **62**, 287-304.

CULTURAL AND PSYCHOLOGICAL DIFFERENCES AMONG OCCUPATIONAL GROUPS*[1]

U. S. Navy Medical Neuropsychiatric Research Unit, San Diego, California

E. K. ERIC GUNDERSON AND JACK L. MAHAN

A. INTRODUCTION

Social background and personality variables are commonly regarded as important factors in choosing or being selected for an occupation. Rosenberg (14) gathered evidence for this proposition in a college student population. Functioning in a specific occupational role has been assumed further to influence or shape personality, particularly in the domain of attitudes and values. One investigator reported negative findings on this question, however (10).

Vroom (17) has reviewed the evidence for social background influences on occupational choice and states that such factors as father's occupation, economic position of the family, and quality of relationships with parents undoubtedly have some effect upon an individual's vocational aspirations and decisions; however, reliable quantitative evidence concerning specific relationships is limited. Furthermore, Vroom did not include demographic and social background variables in his theoretical model for prediction of occupational choice and concluded that the bases for such relationships are still speculative.

In a previous study (9) evidence was developed that differences in personality orientations existed among Navy enlisted occupational groups, roughly dichotomized into "white-collar" and "blue-collar" categories. Another study indicated that Navy occupation was related to parents' socioeconomic status (8). These preliminary investigations suggested the presence of meaningful consistencies between cultural background and personality correlates of occupational role. The present investigation examines cultural and psychological differences among a wider selection of occupational groups and over a greater range of cultural and psychological attributes than previously and relates these differences to actual work and social roles.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on March 10, 1966, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Report Number 66-12, supported by the Bureau of Medicine and Surgery, Navy Department, under Research Work Unit MF 022.01.03-9001. The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Navy. The assistance of Mr. F. A. Thompson, Mr. E. J. Rathswohl, and Mr. J. H. Wineman is gratefully acknowledged.

In the first part of the study, Navy and civilian volunteers for the U. S. Antarctic Research Program were grouped into nine occupational categories and were compared on a number of cultural background characteristics. In the second phase, a new sample of Antarctic volunteers was grouped into the same occupational categories, and differences among groups were evaluated for 47 test and rating variables which reflected a broad sampling of psychological attributes. It was anticipated that those occupational groups which differed most in cultural attributes would differ most on the personality measures. Certain value and personality differences were expected to relate to differences in family socioeconomic status, while others were expected to be specific to the functional requirements or interest patterns of particular occupations. The proposition was tested that value differences among occupations would increase with amount of experience in occupation. Finally, cultural and personality differences among occupational groups were evaluated in terms of effects upon the predictive validities of biographical and personality data and in terms of differences in social and work roles at Antarctic stations.

B. METHOD

1. *Subjects*

Subjects for the first part of the study were 1100 Navy personnel (157 officers and 943 enlisted men) and 305 civilian scientists and technicians who had volunteered for the U. S. Antarctic Research Program (Operation Deep Freeze). Applicants for six expeditions were included in this sample. The mean age of the Navy men was 28 years, and the mean naval experience was 8 years. The mean age of the civilian volunteers, principally physical scientists and meteorologists, was 29 years. Experience levels ranged from graduate student to senior research scientist from university, government, or industrial laboratory.

Subjects for the second part of the study were 583 Navy and 134 civilian volunteers for two Antarctic expeditions. The mean ages for these Navy and civilian samples were 26 and 28 years, respectively.

Characteristics of Antarctic volunteers have been described in detail in a previous publication (4). Navy Antarctic volunteers are above average in intelligence, education, and military performance compared with Navy men generally.

2. *Procedures*

In the first sample each applicant filled out a biographical questionnaire (Standard Psychodiagnostic Record Booklet—Personal History) pertaining to developmental history, including birth place and residence, religion and worship, parental and family background, educational achievement and school adjustment, sports and social interests, medical history and symptoms, and occupational and military experience. Data were coded and transferred to IBM cards for machine tabulations.

Subjects in the second sample filled out a number of personality inventories, including the FIRO-B Inventory, Survey of Interpersonal Values, the Allport-Vernon-Lindzey Study of Values, and two special inventories constructed by the senior author, called the Opinion Survey and the Friend Description Inventories. In both phases of the study, the subjects were exposed to similar psychiatric assessment procedures, but only in the second phase was an extensive personality test battery administered. Part of the applicant population during one year was not given the Survey of Interpersonal Values (SIV), and the sample size is reduced for that portion of the analysis concerned with the SIV scales.

Administrative: Storekeeper, Yeoman, Personnelman, Disbursing Clerk, Aviation Storekeeper, Hospitalman, and Dental Technician.
Technical: Radioman, Electronics Technician, Aerographer, Sonarman, Radarman, and Surveyor.
Mechanical: Machinist Mate, Engineman, Aviation Machinist, Construction Mechanic, and Equipment Operator.
Construction: Builder, Utilitiesman, Steelworker, and Shipfitter.
Electrical: Electricians Mate, Aviation Electrician, and Construction Electrician.
Cooks: Commissaryman.

Three additional occupational groups were formed, one consisting of all naval officer volunteers, another of scientists, and a third of meteorologists and civilian technicians. The principal scientific specialties represented in the scientist group were ionospheric physics, aurora, radio science, geomagnetism, and seismology.

In order to test the hypothesis that the degree of personality differences would increase with amount of occupational experience, five of the Navy enlisted occupational groups were divided by years of naval service into two subgroups, inexperienced (less than eight years) and experienced (eight years or more). One occupational group, the cooks, consisted entirely of experienced men. Differences in means between experienced and inexperienced subgroups within the various occupational groups were tested for the 18 value scales utilized in the study (Study of Values, SIV, and FIRO-B), and differences among the experienced subgroups across occupational groups also were tested.

3. *Personality Inventories*

The Allport-Vernon-Lindzey Study of Values test measures six value orientations based upon a personality classification proposed by Spranger (1). The scales are Theoretical (empirical, rational, critical, intellectual); Economic ("typical American businessman"); Social (altruist and philanthropist); Aesthetic (individualist); Political (oriented toward personal power and influence); and Religious.

The Survey of Interpersonal Values (SIV) purports to measure basic motivational patterns or values (3). The scales, with brief definitions, are Support (being treated with understanding, kindness, and consideration); Conformity (following regulations and doing what is socially correct); Recognition (attracting favorable notice, being looked up to and admired), Independence (doing what one wants to, making own decisions, doing things in one's own way); Benevolence (helping others, being generous); and Leadership (having authority over others, being in a position of leadership).

The FIRO-B questionnaire purports to measure two aspects (expressing and wanting) of three universal interpersonal needs (15). The six scales may be characterized as follows: Expressed Inclusion (participating in group activities); Wanted Inclusion (wanting to be included in group activities); Expressed Control (controlling others, exercising dominance and leadership); Wanted Control (wanting direction or regulation from others); Expressed Affection (being affectionate with others); and Wanted Affection (wanting affection from others).

The remaining test scales and rating measures utilized in the study were developed especially for the Antarctic screening program. The factor analytic technique was employed to identify highly intercorrelated clusters of inventory items which appeared to represent meaningful psychological concepts. Four of the test scales measured common psychological needs: Achievement, Autonomy, Nurturance, and Orderliness. The contents of these four scales are generally similar to those of the corresponding Edwards Personal Preference Schedule Scales (2), although the format of the items was entirely different. Statements were presented as single items to be evaluated on six-point scales ("strongly agree" to "strongly disagree"), and response values were summed for highly intercorrelated items.

Five scales measured attitude toward job and toward the Antarctic assignment. These were called Usefulness, Boredom, Confidence in Organization, Confidence in Medical Care, and Motivation.

Four scales represented personality self-descriptions based upon clusters of intercorrelated traits: Decisive (Decisive, Obedient, Handy, Alert, Orderly, and Self-improving); Excitable (Excitable, Argumentative, Hostile, Suspicious, and Impulsive); Blunt (Blunt, Stubborn, Rough in Manner, and Hard); and Absentminded (Absentminded, Awkward, and Slow).

Seven scales were derived from an inventory consisting of 50 trait adjectives which were rated by subjects on eight-point scales in terms of how much the individual would like those characteristics in a close friend. The scales were labeled Efficiency, Sympathy, Conservatism, Optimism, Dignity, Cynicism, and Humor.

Finally, each candidate was rated independently by psychologists and psychiatrists on a series of personality traits. The two ratings on each trait were summed for each subject to obtain a trait score; differences among the occupational groups in mean trait scores were determined. These data provided comparisons of occupational groups on a variety of personality attributes as viewed by clinical experts. The reliabilities and validities of these trait ratings have been reported elsewhere (5, 6).

4. *Statistical Analyses*

Because of the categorical nature of the personal history data, comparisons were made in terms of differences in percentages falling into specific categories. The chi-square statistic was utilized to test for the significance of differences among groups; because variation on all of the cultural characteristics shown in the results was significant beyond the .05 level of confidence, the χ^2 values and probability levels were not shown for the separate items.

The analysis of variance technique was utilized to test for significance of group differences on all personality measures. *F* ratios were computed based upon differences over all nine of the occupational groups. For those variables with significant *F* ratios ($p < .05$), detailed results in the form of means and standard deviations or deviations from the grand mean were presented for five of the occupational groups.

C. RESULTS

1. *Cultural Characteristics*

Variations in selected cultural characteristics over the nine occupational groups are shown in Table 1. Data for a number of other personal history characteristics—such as religion, frequency of worship, sports interests, symptoms, truancy and delinquency, and parents' marital status—were examined in preliminary stages of the study. Several of these items showed significant differences between total civilian and total Navy groups, but generally did not differentiate within Navy or civilian subgroups. The items shown in Table 1 were considered most relevant to the primary aims of the study, which were to determine the salient cultural differences present and to relate these differences to variance on personality measures. The personal history variables that best discriminated occupational groups in this study—that is, own education, parents' education, and urban-rural residence—were the same items that differentiated occupational groups most sharply in the 10- to 15-year follow-up study of aviation cadets by Thorndike and Hagen (16, p. 40).

Father's occupation was highly related to the man's own occupation. Scientists most often had fathers who were in white-collar occupations (professional, technical, and managerial) and cooks least often—59 per cent and 11 per cent, respectively. Officers' fathers also frequently were in white-collar jobs (56 per cent). Meteorologists' fathers were more frequently white-collar than were those of men in Navy enlisted occupations, and fathers of men in Navy administrative and technical jobs were more frequently white-collar than were those of men in Navy blue-collar trades (Mechanical, Construction, and Electrical groups). Fathers of scientists, civilian technicians, and officers less frequently were in skilled trades than were fathers of men in Navy enlisted occupations. Electricians were an exception among the Navy enlisted occupations; their fathers worked in farming, marine, or forestry jobs relatively frequently (44 per cent) and worked in skilled trades less frequently than any group. Men in Navy construction jobs (carpentry and plumbing) had a relatively high proportion (62 per cent) of fathers in skilled trades. Scientists' and officers' fathers least often worked in farming or forestry.

The results clearly indicate that father's occupation has an influence upon choice of or selection for an occupation far beyond chance. The most striking relationships are the high proportions of scientists and officers who had white-collar fathers and the high proportion of Navy men in skilled trades (Construction) whose fathers were in skilled trades.

Parents' education varied considerably among the nine occupational groups. Parents of the scientist and officer groups had completed high school much more frequently than those of other groups. Parents of civilian technicians had completed high school more frequently than the parents of men in Navy administrative and technical occupations. The pattern for parents' education closely parallels that for father's occupation.

TABLE 1
DIFFERENCES AMONG OCCUPATIONS IN CULTURAL CHARACTERISTICS

	Civilian		Navy						
Variable	Scien.	Tech.	Off.	Admin.	Tech.	Mech.	Constr.	Elec.	Cook
Father's Occupation									
% White-Collar	59	37	56	26	28	20	13	21	11
% Skilled	22	29	23	42	41	39	62	18	44
% Farm, Forest	8	16	10	15	12	23	14	44	22
Father's Education									
% No H.S.	23	34	24	45	40	56	48	48	65
% H.S. Grad.	62	47	58	26	38	25	23	31	22
Mother's Education									
% No H.S.	18	24	19	31	25	37	37	34	40
% H.S. Grad.	68	56	67	46	49	37	32	35	40
Urban-Rural Residence									
% Rural	17	23	22	31	32	46	33	36	33
% Large City	43	30	42	30	34	26	37	28	34
Region of Residence									
% Northeast	32	38	43	32	29	26	38	30	24
% South	16	15	21	25	21	33	24	36	38
Marital Status									
% Married	34	52	58	45	36	59	62	47	56
% Deviation	—32	—19	—19	—11	— 5	— 2	+ 7	+ 1	—24
Education									
% College	99	65	89	20	16	07	07	14	05
Hobbies									
% More than 1	67	56	56	30	38	27	19	25	13
Magazines Read									
% More than 3	38	30	45	17	26	18	17	11	20
N	170	135	157	120	272	273	123	100	55

Moderate differences among groups in urban-rural residence were present. Scientists and officers more often came from large cities and less often from rural areas than did other groups. Mechanics more often than other groups resided in rural areas (46 per cent) and less often resided in large cities. Among the Navy enlisted groups, construction workers (carpenters and plumbers) most often came from large cities (37 per cent), presumably where their fathers would have opportunities to work in skilled trades.

Occupational groups varied by region of residence, but these differences were not large. Navy enlisted occupations, with the exception of construction jobs, tended to have resided in the southern states more frequently and in the northeastern United States less frequently than had other occupational groups. More than one-third of Navy mechanics, electricians, and cooks were drawn from the southern United States, while a quarter or less of other groups were Southerners. Disproportionately few members of the civilian occupations were from the southern United States.

Large variations in marital status were apparent among the nine occupational groups. However, raw differences in percentages married were not meaningful because of age differences among the occupational groups. The relationship of age to percentage married was determined for all Navy volunteers by plotting percentage married against age. From this normative curve, an expected percentage married for each occupational group in terms of its mean age was determined, and this expected percentage married was subtracted from the actual percentage married in each group to provide a measure of deviation from the marriage norm. These values, shown in Table 1, reflect the relative frequency of marriage among members of the various occupational groups with the factor of age taken into account.

By the method described above, scientists were found to be well below their expected marriage rate, while Navy men in Construction, Electrical, and Mechanical groups were at or slightly above the norm. Officers and civilian technicians were much below the marriage norm, while men in enlisted administrative jobs had a marriage rate somewhat below men in blue-collar Navy jobs. Cooks were disproportionately low in the number married, when correction was made for age.

Virtually all of the scientists and a large proportion of the officers had attended college. Relatively few of the Navy enlisted men had attended college; those in administrative and technical jobs more often had had some college experience than had those in other Navy occupations.

Expressed interest in hobbies varied considerably by occupation. Scientists, civilian technicians, and officers frequently reported two or more hobbies, while the Navy enlisted groups much less frequently did so. Men in the Administrative and Technical groups more often had two or more hobbies than did other Navy enlisted occupations, although these differences tended to be small.

Number of magazines read varied moderately with occupation. Officers and scientists most often reported reading more than three magazines regularly, while electricians least often reported reading three magazines. The civilian technicians and Navy technical groups read more than the Navy blue-collar trades, but these differences were small.

Distinctive patterns of cultural characteristics emerged for the major occupational categories. Extreme differences were frequently found between the scientists and officers on the one hand and Navy enlisted blue-collar occupations and cooks on the other. Differences among the Navy enlisted occupations tended to be moderate to slight, but differences present consistently favored the Technical and Administrative groups.

The Scientist and Officer groups did not differ significantly from each other on any of the cultural attributes studied. Both of these groups differed moderately but consistently from the civilian technicians on father's occupation, parents' education, urban-rural residence, and number of magazines read; only on the college attendance item was the difference between the civilian technicians and the Scientist and Officer groups relatively large.

2. *Psychological Attributes*

Results for the psychological data are presented in Tables 2 and 3. The F ratios indicated are for differences among the means for all nine occupational groups. For simplicity of exposition, test results are given for only five groups in Tables 2 and 3. Results for the Meteorologist-Technician group for the most part paralleled those for scientists, although means were consistently less extreme; results for the Administrative group differed little from those for the Technical group, and results for the Construction and Electrical groups as a rule differed negligibly from those for the Mechanical group. The range and nature of the psychological differences present, therefore, are well represented by the five groups included in the tables.

Differences among groups were significant beyond the .01 probability level for all of the Study of Values scales. Differences on the Theoretical scale were very large, mean scores ranging from 51.9 for scientists to 41.9 for cooks, a difference of almost two standard deviations. The Navy enlisted personnel in technical occupations scored next highest (46.6). These results appear highly consistent with occupational requirements.

TABLE 2
VALUE DIFFERENCES AMONG OCCUPATIONAL GROUPS

			Scientists		Officers		Technical		Mechanical		Cooks	
Scale	*F* ratio[a]	*df*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Study of Values												
Theoretical	13.8	8/654	51.9	6.0	45.3	6.4	46.6	6.3	44.2	6.0	41.9	4.7
Economic	6.1	8/654	38.0	8.5	41.0	8.2	42.5	6.2	44.8	6.8	43.4	6.4
Aesthetic	10.3	8/654	39.7	9.5	33.5	8.6	31.9	7.2	30.6	5.4	33.0	7.2
Social	3.9	8/654	32.9	6.0	34.6	7.6	36.4	6.3	37.2	6.4	38.0	4.7
Political	4.5	8/654	40.4	6.4	44.4	6.8	42.1	5.8	42.9	5.7	44.1	4.7
Religious	2.7	8/654	37.3	10.8	41.3	9.4	40.5	7.6	40.3	7.6	39.7	5.3
Survey of Interpersonal Values												
Conformity	10.4	8/527	13.3	6.8	18.6	6.5	20.1	6.2	20.8	5.9	23.4	3.8
Independence	4.5	8/527	18.1	7.5	13.4	6.1	12.9	7.0	13.2	6.4	11.5	4.4
Leadership	4.3	8/527	18.4	6.8	18.7	7.0	15.1	6.1	14.0	6.6	12.0	6.4
FIRO-B												
Ie	2.4	8/691	4.7	1.9	5.9	2.2	5.3	2.0	4.5	2.1	4.7	2.0
Ce	9.3	8/691	2.8	2.2	5.6	2.2	2.3	2.0	2.5	2.4	1.9	1.7
Ae	2.2	8/691	3.3	1.8	4.3	2.7	3.3	2.0	2.9	1.7	2.9	2.1
Aw	2.7	8/691	3.3	2.1	4.5	2.9	3.8	2.3	3.3	2.2	4.2	2.1

[a] *F* ratios are for differences among means for nine occupational groups. Means and standard deviations for civilian technicians, administrative, construction, and electrical groups are not shown in this table.

Navy mechanics scored highest on the Economic scale, which represents practical as opposed to theoretical interests; scientists scored low on this variable. On the Aesthetic scale, scientists again scored very high in relation to other groups, while the Navy Mechanical group scored low. Cooks scored highest on the Social scale, approximately one standard deviation above scientists. Officers scored highest on the Political scale, while cooks scored next highest and scientists scored relatively low. Officers scored highest of all groups on the Religious scale and scientists were at the low end of this scale.

TABLE 3
PERSONALITY DIFFERENCES AMONG OCCUPATIONAL GROUPS
(IN STANDARD DEVIATION UNITS)

Scale	*df*	*F* ratio	Scien.	Off.	Tech.	Mech.	Cook
Need							
Achievement	8/708	3.5	.02	.36	.14	—.12	—.78
Autonomy	8/708	2.6	.20	.27	—.08	—.10	—.60
Attitude							
Usefulness	8/708	13.7	.70	.57	.04	—.29	—.20
Boredom	8/708	4.7	—.55	—.43	.06	.23	—.13
Confidence I	8/708	3.5	—.23	—.15	.29	—.06	.11
Confidence II	8/708	4.5	—.35	—.49	.12	.21	.23
Motivation	8/708	3.7	.13	—.32	.25	.03	.06
Self							
Decisive	8/665	2.5	.02	.31	.03	—.10	—.16
Excitable	8/663	2.2	—.33	—.07	.02	.06	—.12
Friend							
Efficiency	8/665	2.6	—.47	.26	.01	—.09	—.33
Sympathy	8/666	2.2	—.17	.33	—.20	.01	.12
Caution	8/666	12.9	—.96	—.58	—.06	.32	.41
Optimism	8/664	9.1	.60	.49	.01	—.16	—.40
Clinical Evaluations							
Friction	8/705	2.0	.04	.10	—.06	—.03	—.38
Flexible	8/705	2.2	.14	.06	.08	—.09	.22
Acts Out	8/705	2.0	—.13	—.03	—.12	.10	—.12
Modest	8/705	2.4	.13	—.17	.10	—.03	—.03
Conforming	8/705	2.3	.25	—.12	.07	—.11	—.13
Alert	8/705	11.0	.58	.31	.19	—.36	—.53
Persevering	8/705	6.6	.52	.15	.18	—.24	—.60
Assertive	8/705	3.3	.38	.24	—.03	—.18	—.50
Tactful	8/705	6.1	.30	.17	.12	—.22	—.07
Orderly	8/705	10.3	.65	.39	.12	—.41	—.24
Paranoid	8/705	2.9	.05	.56	—.05	.00	—.36

In summary, scientists scored highest of all groups on the Theoretical and Aesthetic scales and scored lowest of all groups on the Economic, Social, Political, and Religious scales. Officers scored highest of the groups shown in Table 2 on the Political and Religious scales. The Technical group scored highest of the Navy groups on Theoretical and lowest of the Navy groups on Political. The Mechanical group scored highest on the Economic scale and lowest of all groups on the Theoretical scale.

Differences among occupational groups were significant on three of the six Survey of Interpersonal Values scales. A very large difference was present on the Conformity scale, with cooks at the high extreme and scientists at the other extreme. Conversely, on the Independent scale, scientists were much higher than other groups, while cooks were at the low end of the scale. On the Leadership scale, officers scored highest and cooks scored approximately one standard deviation lower than officers. Scientists also scored relatively high on Leadership.

Differences among groups were significant for four of the six FIRO-B scales. Officers scored highest on Expressed Inclusion and the Navy mechanics scored lowest on this scale. On the Expressed Control scale, officers scored much higher than any other group, and cooks were at the low end of the scale, almost two standard deviations below the officers. Officers again scored highest on Expressed Affection; differences among other groups were quite small. Officers scored highest on Wanted Affection, cooks scored next highest, and Mechanical group scored lowest.

Results on the Survey of Interpersonal Values and the FIRO-B scales may be summarized as follows: Scientists scored highest of the five groups compared in Table 2 on Independence and scored lowest of any of the groups on Conformity. Officers scored highest on Leadership, Expressed Control, Expressed Affection, and Wanted Affection. Cooks scored highest on Conformity and lowest on Independence and Expressed Control.

Because the remaining scales are experimental and norms have not been published, results for the additional personality scales are presented in terms of deviations of groups from the grand mean. The numerical values shown in Table 3 are standard deviation units for each of the five groups. For example, cooks differed most extremely from other groups on the first scale, Achievement Need, their mean falling .78 standard deviations below the grand mean for all groups. On the same scale, officers scored relatively high—.36 standard units above the grand mean—while mechanics scored slightly below the grand mean, and the Technical group slightly above the grand mean. On the Autonomy scale, officers and scientists tended to score high, while cooks scored at the other extreme, on the average .60 below the grand mean.

All of the five attitude scales were found to discriminate significantly among the nine occupational groups. The large differences among the five groups compared in Table 3 for the Usefulness scale show a range in deviation from the group mean of a .70 deviation for scientists to —.29 for the mechanical group. Officers scored relatively high on feelings of importance of job, .57 standard deviations above the grand mean for all groups. Conversely, results on the Boredom scale indicate that the Mechanical group felt most bored with their past duty assignments, while scientists and officers were least bored.

The Technical group was most confident that the expedition would be well organized, while the scientists were least confident of organizational efficiency. Cooks had most confidence in the medical care to be available, and officers were least confident of the medical care. Scientists also had relatively low confidence in medical care to be available on the expedition. The Technical group expressed highest motivation for the Antarctic assignment, while officers expressed lowest motivation for the assignment.

Two of the four self-description scales significantly discriminated among the nine groups. Of the five groups shown in Table 3, officers described themselves as most decisive and cooks described themselves as least decisive. The Mechanical group tended to describe themselves as slightly more excitable than other Navy groups, while the scientists considered themselves not excitable.

Significant differences were found among the nine occupational groups on four of the seven Friend Description scales. The first scale, called Efficiency, was composed of the following items: Efficient, Industrious, Punctual, Ambitious, and Tidy. Officers were relatively high in preferring traits of efficiency in friends; scientists were low on this scale, approximately one-half standard deviation below the grand mean for all nine groups. Cooks were also relatively low in valuing efficiency in friends.

Officers were relatively high in preferring sympathetic traits in friends. The Sympathy scale included Sympathetic, Praising, Sentimental, Confiding, and Warm. The Navy Technical group was relatively low on this scale, as were scientists. Differences were not large on this variable.

Variations among groups were extreme on the next variable, Conservatism. This scale consisted of the following items: Conservative, Thrifty, Commonplace, Apologetic, Cautious, and Formal. The cooks scored highest on this scale, indicating most preference for conservatism and caution in friends, and the Mechanical group too scored relatively high. Scientists scored extremely low on this scale, and officers scored much lower than the Navy enlisted groups.

On the Optimism scale—which was composed of the items Optimistic, Uninhibited, Curious, Frank, and Studious—a wide range of differences was obtained. Cooks scored relatively low on this scale, and scientists scored at the other extreme—one standard deviation above cooks. Officers also scored high on this scale, while the Mechanical group scored below the grand mean for all groups.

The nine occupational groups differed significantly on 11 of the 19 Clinical Evaluation Ratings. Cooks were rated lowest on Interpersonal Aggressiveness (Friction), Alert, Persevering, Assertive, and Paranoid. Cooks were rated highest on Flexibility. The Mechanical group was rated lowest on Flexible, Tactful, and Orderly. The Mechanical group was rated highest of the five groups shown in Table 3 on "Acts Out."

No support was found for the hypothesis that experience in occupation would affect degree of personality differences. There were relatively few significant differences (six out of 90 possible) between experienced and inexperienced subgroups within the various occupations. Consistent trends were evident for only two variables, Independence (lower with experience) and Leadership (higher with experience). In only one group—Mechanical—did these differences reach statistical significance. Also, there were fewer significant differences among experienced subgroups across occupational groups than among inexperienced subgroups, negating the possibility that more variation existed among experienced subgroups. Thus, it was concluded that experience in occupation was not significantly associated with the number and degree of value differences present among naval enlisted occupational groups.

D. DISCUSSION

The results revealed two types of differences among the occupational groups of this study, differences in cultural background and differences in values and personality traits. These two types of differences appeared to be related to some degree, and our discussion will be focused upon the consistencies between the two sets of data and on the implications of the observed cultural and personality differences for work and social relationships between members of these occupational groups.

The cultural similarities and differences obtained can be largely subsumed under the concept of family socioeconomic status as reflected by father's occupation, parents' education, and urban residence. The individual's own education and occupational status level appears to be highly related to family socioeconomic status level. The Scientist and Officer groups clearly came from quite different social backgrounds than did the civilian technician and Navy enlisted groups, but these two groups (Scientist and Officer) did not differ significantly from each other on any of the cultural attributes. However, because the occupations of military officer and scientist are quite different, value and personality differences associated with differences in occupational roles were anticipated.

The largest differences on value measures between scientists and officers appeared to be highly consistent with occupational role requirements. The most salient differences were on the Theoretical, Aesthetic, Political, Conformity, Independence, and Expressed Control scales. Officers showed largest differences from the enlisted groups on two variables, Leadership and Expressed Control, precisely the areas in which officers are expected to be differentiated from enlisted men.

A larger number of value and personality differences was expected between scientists and cooks than between other groups of the study because the aforementioned groups differed most in cultural characteristics. Fewer differences were anticipated for scientists *versus* civilian technicians; officers *versus* enlisted groups; white-collar (Administrative and Technical enlisted groups) *versus* blue-collar (Mechanical, Construction, and Electrical) enlisted groups; and few if any differences for cooks *versus* blue-collar groups or for specific groups within the white-collar or blue-collar classification *versus* each other—for example, Mechanical *versus* Construction groups.

Generally, these expectations were confirmed. On the personality measures shown in Tables 2 and 3, scientists had many more sizable differences in mean scores (one-half standard deviation or greater) when compared with cooks than when compared with civilian technicians or officers. Officers showed fewer sizable differences from the Technical group than from the Mechanical group or cooks. The Mechanical group showed fewer differences from cooks than from the Technical group, and so on. It will be remembered that the Navy enlisted groups generally were similar to each other on cultural attributes; the Administrative and Technical groups, however, differed consistently—though differences were small—from other enlisted groups on father's occupation (white-collar) and parents' education in the direction of higher socioeconomic status. These groups also tended toward more college attendance and more hobbies.

After selection for an Antarctic expedition, personnel are deployed to one of several scientific stations on the Antarctic continent. Groups of scientists, meteorologists, and Navy men—usually including two radiomen, an electronics technician, a hospital corpsman, a cook, and one or more mechanics, plumbers, carpenters, electricians, and equipment operators—operate these stations for an entire year. These groups must be completely self-sustaining because there is no possibility of relief or rescue for several months, from about February to November. Performance data, including sociometric choices, are routinely gathered at these stations, and it is possible to compare occupations on performance criteria and on measures of work and social interaction.

The many differences in cultural characteristics and values among occupational groups participating in the Antarctic program indicate that the validities of these variables for prediction of performance may be attenuated if not destroyed. Studies of the Navy, scientist, and meteorologist-technician participants have indicated that biographical attributes are not equally effective as predictors of performance for these three occupational subgroups (7, 12, 13). The results for the clinical evaluations shown in Table 3 made it clear that ratings by clinicians were positively correlated with socioeconomic status level, scientists generally receiving favorable evaluations and cooks receiving unfavorable ones. The fact that cooks achieve more favorable performance evaluations than do other groups indicates the futility of utilizing clinical ratings to predict performance when all occupational groups are combined. It seems clear that specific variables must be identified which are relevant to performance in each major occupational subgroup separately. Development of different linear regression prediction equations for separate occupational subgroups has tended to confirm this assumption.

Differences among occupational groups in their social roles at Antarctic stations were examined by means of sociometric data. Members of 16 groups gave choices for "close friends" on a sociometric form after a period of several months of group isolation. Cooks had more social relationships than did any other occupational group, as indicated by the average number of friendship choices given and received. Radio and electronics technicians were lowest in the number of friendship relationships, while blue-collar enlisted men (carpenters, plumbers, electricians, etc.) were high in number of friendships. Meteorologists and civilian technicians were moderately sociable in terms of friendship choices, while scientists were relatively low in amount of social relationship.

Similarly, differences were present among occupational groups in amount of interaction with others on the job. When choices given and received on a sociometric item indicating persons with whom one worked were averaged for each occupational group, Construction, Mechanical, and Electrical groups interacted with others on the job more than did radio and electronics technicians or meteorologists. Cooks, due to receiving assistance from all or almost all of the station members in cooking duties, had the highest amount of work interaction; scientists had the lowest amount.

The differences in social and work roles among occupational groups appear consistent with certain differences in values expressed in the psychological tests. The cooks and the Mechanical group scored relatively high on the Social and Political scales, were high in social conformity and conservatism (caution) as indicated by the Survey of Interpersonal Values and Friend Description scales, and preferred traits of sympathy and warmth in friends. Occupational groups more concerned with technical tasks and problems (scientists and Navy Technical group), and presumably less concerned with interpersonal relationships, scored higher than cooks and mechanics on the Theoretical Scale and Achievement Need and lower than the same groups on the Social and Sympathy scales.

Cultural and value differences may have practical importance in forecasting group compatibility in such settings as Antarctic small stations where persons in very dissimilar occupational roles must interact to achieve common goals. If differences are too great, reductions in group harmony and effectiveness might be expected. Some evidence for this hypothesis has been adduced from studies of friendship choices within small work groups (11). Heterogeneity on cultural background characteristics and personality needs tended to be related to incompatibility among work associates during the latter part of the Antarctic winter period under conditions of maximum confinement and isolation from the outside world.

E. SUMMARY

Cultural background and psychological differences were determined for nine occupational groups which are represented in the U. S. Antarctic Research Program. Cultural differences obtained could be largely subsumed under the concept of family socioeconomic status levels. Variability in personality measures tended to be consistent with those for cultural attributes. These occupational differences appear to have had a profound effect upon validities of biographical and personality data for performance prediction. Social and work roles at Antarctic small stations appeared generally consistent with the cultural and personality data. Heterogeneity in cultural backgrounds and personal values may affect the compatibility and effectiveness of work groups in isolated settings.

REFERENCES

1. ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. A Study of Values (rev. ed.). Cambridge, Mass.: Houghton Mifflin, 1951.
2. EDWARDS, A. L. Manual for the Edwards Personal Preference Schedule (rev. ed.). New York: Psychological Corp., 1959.
3. GORDON, L. V. Manual for Survey of Interpersonal Values. Chicago: Science Research Assoc., 1960.

4. Gunderson, E. K. E. Personal and social characteristics of Antarctic volunteers. *J. Soc. Psychol.*, 1964, **64**, 325-332.
5. ———. The reliability of personality ratings under varied assessment conditions. *J. Clin. Psychol.*, 1965, **21**, 161-164.
6. Gunderson, E. K. E., & Kapfer, E. L. The predictive validity of clinical ratings for an extreme environment. *Brit. J. Psychiat.*, in press.
7. Gunderson, E. K. E., & Nelson, P. D. Biographical predictors of performance in an extreme environment. *J. of Psychol.*, 1965, **61**, 59-67.
8. ———. Socioeconomic status and Navy occupations. *Person. & Guid. J.*, 1965, **44**, 263-265.
9. ———. Personality differences among Navy occupational groups. *Person. & Guid. J.*, in press.
10. Miller, S. Relationship of personality to occupation, setting, and function. *J. Counsel. Psychol.*, 1962, **9**, 115-121.
11. Nelson, P. D. Compatibility among work associates in isolated groups. Report Number 64-13, U. S. Navy Medical Neuropsychiatric Research Unit, San Diego, California, 1964.
12. Nelson, P. D., & Gunderson, E. K. E. Personal history correlates of performance among military personnel in small Antarctic stations. Report Number 63-20, U. S. Navy Medical Neuropsychiatric Research Unit, San Diego, California, 1963.
13. Nelson, P. D., & Orvick, J. M. Personal history correlates of performance among civilian personnel in small Antarctic stations. Report Number 64-4, U. S. Navy Medical Neuropsychiatric Research Unit, San Diego, California, 1964.
14. Rosenberg, M. Occupations and Values. Glencoe, Ill.: Free Press, 1957.
15. Schutz, W. C. FIRO: A Three-Dimensional Theory of Interpersonal Behavior. New York: Rinehart, 1958.
16. Thorndike, R. L., & Hagen, E. 10,000 Careers. New York: Wiley, 1959.
17. Vroom, V. H. Work and Motivation. New York: Wiley, 1964.

Special Environments Division
U. S. Navy Medical Neuropsychiatric Research Unit
San Diego, California 92152